Army Science Conference Proceedings

12-15 June 1990

Volume I

Principal Authors A through F

This document has been approved for public release and sale; its distribution is unlimited.

19970929 068

DEPARTMENT OF THE ARMY
OFFICE OF THE ASSISTANT SECRETARY
WASHINGTON, DC 20310-0103

SARD-TR

3 0 JUL 1990

MEMORANDUM FOR SEE DISTRIBUTION

SUBJECT: Proceedings of the 1990 Army Science Conference

The 17th Army Science Conference was held at the Omni Durham Hotel and Convention Center, Durham, North Carolina, 12-15 June 1990. The conference presented a cross section of the many significant scientific and engineering programs carried out by the Department of the Army (DA). Additionally, it provided an opportunity for DA civilian and military scientists and engineers to present the results of their research and development efforts before a distinguished and critical audience.

These Proceedings of the the 1990 Army Science Conference are a compilation of all papers presented at the conference and the supplemental papers that were submitted.

Our purpose for soliciting these papers was to:

a.   Stimulate the involvement of scientific and engineering talent within the Department of the Army;

b.   Demonstrate Army competence in research and development;

c.   Provide a forum wherein Army personnel can demonstrate the full scope and depth of their current projects; and

d.   Promote the interchange of ideas among members of the Army scientific and engineering community.

The information contained in these volumes will be of benefit to those who attended the conference and to others interested in Army research and development. It is requested that these Proceedings be placed in technical libraries where they will be available for reference.

George T. Singley III
Deputy Assistant Secretary
For Research and Technology

DISTRIBUTION:

Office of the Under Secretary of Defense for Research and Advanced Technology, Wash, DC 20310-3080

Office of the Assistant Secretary of the Army (RD&A), ATTN: SARD-TR, Wash, DC 20310-0103

HQDA, Assistant Chief of Staff for Intelligence, ATTN: DAMI-ZA, Wash, DC 20310

HQDA, Deputy Chief of Staff for Logistics, ATTN: DALO-ZX, Wash, DC 20310-0500

HQDA, Deputy Chief of Staff for Operations & Plans, ATTN: DAMO-FD, Wash, DC 20310-0400

HQDA, Director of Information Systems for C4, ATTN: SAIS-ZA, Wash, DC 20310-0400

HQDA, Deputy Chief of Staff for Personnel, ATTN: DAPE-ZA, Wash, DC 20310-0300

Office of the Chief of Engineers, ATTN: CERD-M, 20 Massachusetts Avenue, N.W., Wash, DC 20314-1000

Office of the Surgeon General, ATTN: DASG-ZA, 5109 Leesburg Pike, Falls Church, VA 22041-3258

COMMANDERS/DIRECTORS:

USA Concepts Analysis Agency, 8120 Woodmont Avenue, Bethesda, MD 20814-2797

USA Operational Test & Evaluation Agency, 5600 Columbia Pike, Falls Church, VA 22041-5115

USA Information Systems Engineering Command, Ft. Huachuca, AZ 85613-5000

USA Strategic Defense Command, P. O. Box 15280, Arlington, VA 22215-0150

USA Materiel Command, ATTN: AMCLD, 5001 Eisenhower Avenue, Alexandria, VA 22333-0001

  USA Materiel Systems Analysis Activity, ATTN: AMXSY-D, APG, MD 21005-5071

  USA Armament, Munitions & Chemical Command, ATTN: AMSMC-CG, Rock Island, IL 61299-6000

  USA Armament RD&E Center, ATTN: SMCAR-TD, Picatinny Arsenal, NJ 07806-5000

    Fire Support Armament Cen, ATTN: SMCAR-FS, PA, NJ 07806-5000

    Armament Engineering Dir, ATTN: SMCAR-AE, PA, NJ 07806-5000

    Close Combat Armament Cen, ATTN: SMCAR-CC, PA, NJ 07806-5000

    Chemical RD&E Center, ATTN: SMCCR-TD, APG, MD 21010-5423

USA Aviation Systems CMD, ATTN: AMSAV-GTD, 4300 Goodfellow Blvd, St. Louis, MO 63120-1798

USARTA, Ames Research Center, ATTN: SAVRT-D, Moffett Field, CA 94035-1099

USARTA, Aeroflightdynamics Dir, ATTN: SAVRT-AF, Moffett Field, CA 94035

USARTA, Aviation Applied Technology Dir, ATTN: SAVRT-TY, Ft Eustis, VA 23604

USARTA, Propulsion Dir, ATTN: SAVRT-PN, Lewis Research Center, 21000 Brook Park Rd, Cleveland, OH 44135

USARTA, Aerostructures Dir, ATTN: SAVRT-SD, Hampton, VA 22665

  Avionics R&D Activity, ATTN: SAVAA, Ft Monmouth, NJ 07703-5000

USA Communications-Electronics Cmd, ATTN: AMSEL-TDD, Ft Monmouth, NJ 07703-5001

  Center for Command, Control & Communications System, ATTN: AMSEL-RD-C3, Ft Monmouth, NJ 07703-5202

  Center for Night Vision & Electro-Optics, ATTN: AMSEL-RD-NV-D, Ft Belvoir, VA 22060-5677

  Electronics Warfare/RSTA Center, ATTN: AMSEL-RD-EW-D, Ft Monmouth, NJ 07703-5303

  Signals Warfare Lab, ATTN: AMSEL-RD-SW-D, Vint Hill Farms Station, Warrenton, VA 22186-5100

USA Laboratory Command, ATTN: AMSLC-TD, 2800 Powder Mill Road, Adelphi, MD
   29783-1145
   Atmospheric Sciences Lab, ATTN: SLCAS-DD, WSMR, NM 88002-5501
   Ballistic Research Lab, ATTN: SLCBR-OD, APG, MD 21005
   Electronics Technology & Devices Lab, ATTN: SLCET-D, Ft Monmouth, NJ
      07703-5000
   Electronic Warfare Vulnerability Assessment Lab, ATTN: SLCEW-M-D, WSMR, NM
      8802-5513
   Harry Diamond Labs, ATTN: SLCHD-D, Adelphi, MD 20783-1197
   Human Engineering Lab, ATTN: SLCHE-H, APG, MD 21005-5001
   Materials Technology Lab, ATTN: SLCMT-D, Watertown, MA 02172-0001
   Army Research Office, ATTN: SLCRO-D, RTP, NC 27709-2211
   Army Research, Development & Standardization Group (Europe) Box 65, FPO
      New York 09510
USA Missile Command, ATTN: AMSMI-R, Redstone Arsenal, AL 35898-5240
USA Tank-Automotive Cmd, ATTN: AMSTA-CG, Warren, MI 48397-5000
   Aberdeen Proving Ground, ATTN: STEAP-OC, APG, MD 21005-5001
   Dugway Proving Ground, ATTN: STEDP-CO, Dugway, UT 84022-5000
   Electronic Proving Ground, ATTN: STEEP-DO, Ft Huachuca, AZ 85613-7110
   Aviation Development Test Activity, ATTN: STEBG-CO, Ft Rucker, AL 36362-5276
   Combat Systems Test Act, ATTN: STECS-CO, APG, MD 21005-5059
   White Sands Missile Range, ATTN: STEWS-CG, WSMR, NM 88002-5000
USA Aviation Systems Cmd, ATTN: AMSAV-G, 4300 Goodfellow Blvd, St Louis, MO
   63120-1798
   Belvoir RD&E Center, ATTN: STRBE-Z, Ft. Belvoir, VA 21005-5606
   Natick RD&E Center, ATTN: STRNC-Z, Natick, MA 01760-5000
US Army Corps of Engineers
   Cold Regions Rsch & Eng Lab, ATTN: CRREL-TD, P. O. Box 282, Hanover, NH
      03755-1290
   Construction Eng Rsch Lab, ATTN: CERL-Z, P. O. Box 4005, Champaign, IL
      61820-135
   Engineer Topographic Labs, ATTN: ETL-TD, Ft Belvoir, VA 22060-5546
   Waterways Experiment Station, ATTN: WESTV, P. O. Box 631, Vicksburg, MS
      39180-0631
USA Medical R&D Command, ATTN: SGRD-ZA, Ft Detrick, Frederick, MD 21701-5012
   Aeromedical Rsch Lab, ATTN: SGRD-UAC, Ft Rucker, AL 36362-5292
   Inst of Dental Rsch, WRAMC, ATTN:SGRD-UDZ, Wash, DC 20307-5300
   Inst of Surgical Rsch, ATTN: SGRD-USZ, Ft Sam Houston, TX 78234-6200
   Letterman Army Inst of Rsch, ATTN: SGRD-ULZ, Presidio of San Francisco, CA
      94129-6800
   Biomedical R&D Lab, ATTN: SGRD-UBZ, Frederick, MD 20701-5010
Medical Rsch Inst of Chemical Defense, ATTN: SGRD-UV-ZA, APG, MD 21010-5425
Medical Rsch Inst of Environmental Medicine, ATTN: SGRD-UEZ, Natick, MA
   01760-5007
Medical Rsch Inst of Infectious Diseases, ATTN: SGRD-UIZ-A, Ft Detrick,
   Frederick, MD 21701-5011
Walter Reed Army Inst of Rsch, ATTN: SGRD-UWZ, Washington, DC 20307-5100
USA Health Services Command, Ft. Sam Houston, TX 78234-61000
USA Environmental Hygiene Agency, ATTN: HSHB-Z, APG, MD 21010-5422

USA Research Institute for the Behavioral and Social Sciences, ATTN: PERI-ZT, 5001 Eisenhower Ave, Alexandria, VA 22333-5600
  ARI Field Unit, ATTN: PERI-IJ, POB 2086, Ft Benning, GA 31905
  ARI Field Unit, ATTN: PERI-SB, POB 6057, Ft Bliss, TX 79906
  ARI Field Unit, ATTN: PERI-SH, Ft Hood, TX 76544-5065
  ARI Field Unit, ATTN: PERI-IK, Steele Hall, Ft Knox, KY 40121
  ARI Field Unit, ATTN: PERI-SL, POB 290, Ft Leavenworth, KS 66027
  ARI Field Unit, ATTN: PERI-IR, Ft Rucker, AL 36362-5000
  ARI Field Unit, ATTN: PERI-IO, POB 5787, Presidio of Monterey, CA 93940
USA Training and Doctrine Command, ATTN: ATDO-ZT, Ft Monroe, VA 23651-5000
  Combined Army Test Act, ATTN: ATCT-CG, Ft Hood, TX 76544-5065
COMMANDANTS:
USA Air Defense Cen, ATTN: ATZC-CG, Ft Bliss, TX 79916-5000
USA Armor Cen, ATTN: ATZK-CG, Ft Knox, KY 40121-5000
USA Aviation Cen, ATTN: ATZQ-CG, Ft Rucker, AL 36362-5000
USA Chemical Cen, ATTN: ATZN-CG, Ft McClellan, AL 36205-5000
USA Combined Arms Cen, ATTN: ATZL-SC, Ft Leavenworth, KS 66027-5000
USA Engineer Cen, ATTN: ATZA-CG, Ft Belvoir, VA 22060-5000
USA Infantry Cen, ATTN: ATZB-CG, Ft Benning, GA 31905-5000
USA Intelligence Cen, ATTN: ATSI-SA, Ft Hauchuca, AZ 85613-7000
USA Ordnance Cen, ATTN: ATSL-CMT, APG, MD 21005-5201
USA Signal Cen, ATTN: ATZH-CH, Ft Gordon, GA 30905-5000
USA Field Artillery Center, ATTN: ATZR-C, Ft Sill, OK 73503
SUPERINTENDENT:
US Military Academy, ATTN: Technical Library, West Point, NY 10996
COPIES FURNISHED:
Defense Advanced Research Projects Agency, 1400 Wilson Blvd, Arlington, VA
  22209-2308
Defense Logistics Agency, Cameron Station, Alexandria, VA 22304-6183
Defense Technical Information Center, Cameron Station, Alexandria, VA
  22304-6145
Lawrence Livermore National Lab, ATTN: L-191, POB 808, Livermore, CA 94550
Los Alamos National Lab, ATTN: Dir for Energy, Rsch & Tech, Los Alamos, NM
  87545
NASA HQS, Suite 4237, 400 Maryland Ave, SW, Wash, DC 20546
National Science Foundation, 1800 G Street, NW, Wash, DC 20550
United Nations Library, ATTN: Acquisition Section, Room L-138A, New York, NY
  10017
US NAVY
  Naval Air Systems Command, Code AIR-03-D, 1411 Jefferson Davis Hwy,
    Arlington, VA 22202-3000
  Naval Rsch Lab, ATTN: Dir of Rsch, Washington, DC 20375
  Office of Naval Rsch, Code 10, 800 North Quincy St, Arlington, VA 22217-5000
  HQ US Marine Corps, Code RD-1, Washington, DC 29380-0001
US AIR FORCE
  Air Force Systems Command, ATTN: Technical Director, Andrews AFB, Washington,
    DC 20334-5000
  Air Force Office of Scientific Research, ATTN: Tech Director, Bolling AFB,
  Washington, DC 20332-6448

PROCEEDINGS

OF THE

1990 ARMY SCIENCE CONFERENCE

DURHAM, NORTH CAROLINA

12-15 JUNE 1990

VOLUME I

Principal Authors A through F

TABLE OF CONTENTS

PROCEEDINGS OF THE 1990 ARMY SCIENCE CONFERENCE

i

AUTHOR	TITLE	VOL	PAGE
Doherty, Tammy J.	See Bruttig, Stephen P.	I	187
Dominessy, Mary E.	See Lukas, Jeffrey H.	II	423
Dornath-Mohr, Michelle A.	Shallow, Low Temperature Au-Ge Ohmic Contacts to GaAs	I	475
Dubois, Doria R.	See Bancroft, William H.	I	73
Dulaney, Marland D., Jr.	Identification of a New Cyanide Countermeasure	I	491
Duncan, Fred	See Keniston, Richard C.	II	281
Dunne, C. P.	See Porter, William L.	III	93
Dutta, Mitra	Semiconductor Optical Waveguide Devices	I	507
Dutta, Mitra	See Choi, K. K.	I	311
Dutta, Mitra	See Cole, Melanie W.	I	349
Dutta, Mitra	See Smith, Doran D.	III	341
Eckart, Donald W.	See Dornath-Mohr, Michelle A.	I	475
Eckels, Kenneth	See Bancroft, William H.	I	73
Enriquez, John I., Sr.	See Keniston, Richard C.	II	281
Evans, Christopher	Metabolic Rescue with a Precursor for Adenylate Synthesis Stimulates Phosphate Utilization and Prolongs Survial in Hemorrhagic Shock	I	517
Falls, Terril C.	Automated Analysis of Interactions Between Terrain and Ground Force Mobility	I	533
Farrand, Timothy G.	See Magness, Lee S.	II	465
Fazi, Christian	Breakdown in Silicon p-n Junctions Due to Short Pulse, High-Field Electromagnetic Stress	I	543
Fedele, Paul D.	Aerosol Transport Modeling Indicates Hazards and Improved Protective Measures for Dusty Agents	IV	1
Ferry, Michael	See Daunt, Geraldine	I	451
Fleetwood, R.	See Bushell, M.	I	227

AUTHOR	TITLE	VOL	PAGE
Gallo, Benedict J.	See Steeves, Diane M.	III	395
Garvin, Charles	Hybrid Approaches for Spread-Spectrum COMINT	II	27
Garvin, Charles	See Brody, Philip S.	I	159
Goff, John R.	Optical Associative Memory Based on Binary Phase Algebra and the Inner-Product Architecture	II	39
Gold, Kenneth	Biodegradation of Energetic Materials	II	53
Gordon, Daniel	See Sadoff, Jerald C.	III	153
Gordon, Richard K.	See Smejkal, Ruthann M.	III	329
Gowenlock, David A.	See Gallo, Benedict J.	II	17
Graham, Scott E.	Soldier Performance Research Project: Armor Field and SIMNET Tests	II	65
Grant, D. F.	See Porter, William L.	III	93
Graves, Bruce R.	See Mullins, William D.	IV	65
Green, Shawn J.	Identification of a Novel Cytokine-induced Effector Molecule Involved in the Intracellular Destruction of Leishmania, Franciscella and Plasmodium Sporozites: Nitric Oxide	II	81
Grillo, F. Gregory	Protection from Ricin Toxicity by Brefeldin A	II	95
Grisham, John A.	See Tanton, George A.	III	447
Grogl, Max	See Milhous, Wilbur K.	II	543
Gross, Clark L.	See Smith, William J.	III	357
Grower, Marvin F.	Development of Topically Applied Anti-Inflammatory Agents for Treating Acute Pulpal Inflammation in Combat Situations	II	103
Guidos, Bernard J.	Supercomputers Simulation of Supersonic Viscous Flow Over Pointed, Spherical, and Flat Tipped Shell	II	119
Gupta, N.	See Simonis, G. J.	III	287

AUTHOR	TITLE	VOL	PAGE
Howard, Stephen L.	See Sausa, Rosario C.	III	209
Hoyt, Reed W.	Evaluating Combat Soldier Energy Balance in the Field with Stable Isotope and Ambulatory Monitoring Methods	II	193
Hsieh, Alex J.	Solvent Stress Cracking and Failure Mechanisms in Polyetherimide Composites	II	207
Hubbard, Roger W.	See Armstrong, Lawrence E.	I	59
Hubbard, Roger W.	See Szlyk, Patricia C.	III	437
Hudson, Thomas	See Grillo, F. Gregory	II	95
Ingram, James K.	See Franco, Raphael A.	I	617
Intaglietta, Marcos	See Bruttig, Stephen P.	I	187
Jackson, Ronald	See Darrigrand, Andre A.	I	441
Jaglowski, A. J.	See Singler, R. E.	III	315
Johnson, John L.	Object Extraction for Automatic Target Recognition	II	223
Johnson, John L.	See Anitole, George	I	45
Johnson, Robert A.	Laser-Induced Synthesis of Methoxymethanol	II	253
Johnson, Ronald L.	See Anitole, George	I	45
Johnson, Thomas H.	Xenon Chloride Laser Scaling	II	239
Jones, Bruce H.	See Friedl, Karl E.	I	631
Jones, Kenneth A.	See Dornath-Mohr, Michelle A.	I	475
Jones, Tanya E.	See Hoyt, Reed W.	II	193
Jordan, Debbee B.	See Mullins, William D.	IV	65
Kalathil, Biju	See Crawford, Kenneth H.	I	389
Kaste, Richard C.	An Experimental Artillery Division Aid	IV	49
Kearl, Cyril E.	How Much Soldier Quality? Estimating Cost-Effective Recruit Selection Policy for a Smaller Army	II	265

AUTHOR	TITLE	VOL	PAGE
Milhous, Wilbur K.	Circumventing Antimalarial Drug Resistance: Theoretical and Clinical Implications	II	543
Miller, Mary J.	See Sharp, Edward J.	III	263
Millnamow, Gregory A.	See Bruttig, Stephen P.	I	187
Miziolek, Andrezej W.	Novel Laser-Based Detector for Gas Chromatography	II	555
Miziolek, Andrzej W.	See Forch, Brad E.	I	573
Miziolek, Andrzej W.	See Forch, Brad E.	I	587
Miziolek, Andrzej W.	See Forch, Brad E.	I	603
Miziolek, Andrzej W.	See McQuaid, Michael J.	II	507
Miziolek, Andrzej W.	See Sausa, Rosario C.	III	209
Miziolek, Andrzej W.	See Sausa, Rosario C.	III	225
Monty, Richard A.	See Lukas, Jeffrey H.	II	423
Moore, Gerry L.	Additive Solutions for the 21 Day Preservation of Previously Frozen Red Blood Cells	II	565
Morris, Jeffrey B.	See Forch, Brad E.	I	603
Morris, Jeffrey B.	See Miziolek, Andrzej W.	II	555
Morrison, Walter F.	A Modified Lagrange Pressure Gradient for the Regenerative Liquid Propellant Gun	II	577
Morrison, Walter F.	See Coffee, Terence P.	I	323
Mullins, William D.	Test and Evaluation of Man Portable Lasers for Infantry Use IV	IV	65
Murphy, Charles H.	Side Moment Exerted by a Spinning, Coning, Highly Viscous Liquid Payload	II	593
Nacy, Carol A.	See Green, Shawn J.	II	81
Nattress, D.	See Porter, William L.	III	93
Nelson, Douglas C.	See Burrows, W. Dickinson	I	201
Nelson, Douglas C.	See Fedele, Paul D.	IV	1
Nelson, James H.	See Burrows, W. Dickinson	I	201

AUTHOR	TITLE	VOL	PAGE
Sims, S. Richard F.	An Integrated Target Acquisition System for Fire Control and Autonomous Acquisition	III	303
Singler, R. E.	Poly(organophosphazenes) for Nonlinear Optical Applications	III	315
Sjogren, Maria H.	See Bancroft, William H.	I	73
Smejkal, Ruthann M.	Binary Antidotes for Organophosphate Chemical Warfare Agents	III	329
Smith, Doran D.	The Effects of Electron Beam Processing on High Electron Mobility Transistors	III	341
Smith, Doran D.	A Selectively-Contacted Dual Channel High Electron Mobility Transistor	III	347
Smith, M.	See Bushell, M.	I	227
Smith, William J.	Sulfur Mustard-induced Biochemical Alterations in Proliferating Human Cells in Culture	III	357
Sousk, Stephen F.	Motion Planning for the Universal Self-Deployable Cargo Handler	III	369
Sparenborg, Steven	The Excitatory Amino Acid Antagonist MK-801 Prevents Nerve Agent-induced Neuropathology Even When Given After the Onset of Convulsions	III	383
Stanley, Ann E.	See Johnson, Robert A.	II	253
Steeves, Diane M.	Stereospecificty of Microbial Enzymes for G-Agent Detoxification	III	395
Steinhaus, Ralph K.	Formation of Methemoglobin and Metmyoglobin Using 8-Aminoquinoline Derivatives or Sodium Nitrite and Subsequent Reaction with Cyanide	III	409
Stensby, John	See Tanton, George A.	III	447
Still, G.	See Bushell, M.	I	227
Sturek, Walter B.	See Weinacht, Paul	III	503

AUTHOR	TITLE	VOL	PAGE
Swinson, Mark L.	Military Robotics: A Technology Assessment From a User's Perspective	III	423
Szafraniec, Linda L.	See Yang, Yu-Chu	III	547
Szlyk, Patricia C.	An Innovative Fluid Delivery System for Chemical Protective Clothing: An Evaluation	III	437
Tanton, George A.	UV-IR Detector and Focal Plan Array Material Evaluation Using Faraday Rotation	III	447
Taub, Irvin A.	See Kim, Hie-Joon	II	295
Thompson, Donald L.	See Chabalowski, Cary F.	I	295
Thompson, Thomas J.	See Pleban, Robert J.	III	69
Trevino, S. F.	A Determination of the Intermolecular Potential of Nitromethane	III	461
Underwood, Robert B.	See Falls, Terril C.	I	533
Valdes, James J.	Direction of Toxins with a Reversible Biosensor	III	477
Valentine, Patrick J.	See Pleban, Robert J.	III	69
Vanderhoff, John A.	Simultaneous Determination of Temperatures of OH Concentrations in a solid Propellant Flame	III	491
Vlahocos, Constaine	See Dulaney, Marland D., Jr.	I	491
Vodkin, M. H.	See Hoover, Timothy A.	II	181
Vogel, James A.	See Friedl, Karl E.	I	631
Wade, Charles E.	See Bruttig, Stephen P.	I	187
Walker, Clinton	See Busciglio, Henry H.	I	213
Walker, John E.	See Halliday, John W.	II	135
Walker, John E.	See Steeves, Diane M.	III	395
Walkinshaw, John W.	See Perkins, Janet	III	55

AUTHOR	TITLE	VOL	PAGE
Ward, Alford L.	See Fazi, Christian	I	543
Weinacht, Paul	Navier-Stokes Predictions of Pitch Damping for High L/D Finned Kinetic Energy Projectiles	III	503
Weinacht, Paul	See Guidos, Bernard J.	II	119
Weiner, M.	See Kim, Anderson	II	307
Weiner, Maurice	See Braun, Christopher G.	I	145
Westgate, Roger C.	See Fazi, Christian	I	543
Wheeler, Thomas J.	See Richardson, Judith D.	III	109
White, Leonard A.	Setting Enlistment Standards on the ABLE to Reduce Attrition	III	519
Wiesmann, William	See Evans, Christopher P.	I	517
Williams, Jim C.	See Hoover, Timothy A.	II	181
Willingham, Reginald A.	See Singler, R. E.	III	315
Wisher, Robert A.	Predicting the Decay of Mobile Subscriber Equipment (MSE) Operator Skills	III	533
Wisler, John A.	See Dulaney, Marland D., Jr.	I	491
Wisniewski, Henry L.	See Santiago, Joseph M.	III	179
Wolfe, A. David	See Smejkal, Ruthann M.	III	329
Wood, Gary L.	See Sharp, Edward J.	III	263
Woontner, Susan	See O'Neill, Timothy R.	III	15
Wren, Gloria P.	See Coffee, Terence P.	I	323
Yang, Yu-Chu	Oxidative Decontamination of the Chemical Nerve Agent	III	547
Yip, Pearl W.	See Perkins, Janet	III	55
Youmans, R.	See Kim, Anderson	II	307
Younger, M.	See Garvin, Charles	II	27
Zeto, R.	See Kim, Anderson	II	307

Designer Energetic Materials

*George F. Adams, Dr.
Cary F. Chabalowski, Dr.
Betsy M. Rice, Dr.
Ballistic Research Laboratory, SLCBR-IB-I
Aberdeen Proving Ground, MD   21005-5066

INTRODUCTION: To achieve the goals set in heavy force modernization strategies, virtually all aspects of gun ballistics must be improved. It is particularly important that the interior ballistics performance of the weapons system provide an efficient and reproducible transfer of energy to the bullet. This requirement has led to the development of new, innovative gun propulsion technologies, including liquid propellants, electromagnetic propulsion and electrothermal propulsion. Each of these propulsion concepts attempts to provide efficient, reproducible and safe performance. To achieve these goals an aggressive research and development program to find high energy density materials for liquid and solid propulsion applications, and to find new materials for energy storage has been undertaken. Since the synthesis and testing of new materials is both costly and time consuming, alternative approaches are needed that provide performance related metrics. This contribution describes the application of empirical molecular modeling techniques to the study of high energy density materials.

The ability of scientists to design new materials for specific applications has been amply demonstrated during the past quarter century. Chemical and pharmaceutical companies employ staffs of theoretical chemists, biologists, computer professionals and, in some cases, professional graphics artists whose task is to develop and exercise computational methods in order to provide both evolutionary and revolutionary advances in the design of more effective products. As the cost of computer cycles is reduced, competitive standing requires the use of theoretical chemical methods as a powerful screening tool in the evaluation of proposed new products. This contribution will describe the modification and extension of these standard molecular modeling tools in order that we may evaluate and design new materials for Army applications. In particular, we address the need to develop appropriate interaction terms for those classes of chemical compounds often used in

1

energetic materials applications. These investigations are motivated, in part, by discussions with Army scientists at the 1988 Army Science Conference. These discussions alerted us to the potential utility of the molecular modeling tools, while informing us of the problem that hampered such applications: the lack of appropriate force fields.

In the following section we review the various theoretical chemical methods employed in our research. These range from empirically derived molecular mechanics techniques to state of the art ab initio quantum chemical methods. Two classes of chemical problems are addressed: test problems that establish our ability to reproduce results obtained in other laboratories, and problems dealing with the structure and properties of new energetic molecules. We will describe the techniques we've used to broaden the applicability of the molecular modeling tools to include high energy density compounds. Once the rationale for our research strategy is established, we will provide several examples of the type of information that can be obtained using the methods, and we will comment upon the utility of the techniques in energetic materials research. We will conclude by suggesting the directions the research will take in the near future.

THEORETICAL METHODS: The MM3 force field for hydrocarbons and closely related compounds was described in a series of papers by Allinger, et al., recently.[1-3] The first paper summarized the newly defined force field for hydrocarbons and commented upon the need to add terms that were not included in the MM2 and earlier force fields. The new formalism permits the chemist to predict with good accuracy the structure and heat of formation of a hydrocarbon, as well as the vibrational frequencies of the compound. As was the case for earlier versions of this force field, the parameters were derived by fitting enormous quantities of experimental data. It is this approach to deriving appropriate force field parameters that limits the current applications of this paradigm of molecular modeling. Since there are many classes of chemical compounds that are not well characterized in terms of structure, energy content or, in some cases, spectroscopy, it is not possible to fit MM3 (or MM2) parameters for these classes of chemical species. An alternative to the fitting of empirical data was described by Hoffinger and coworkers,[4] wherein the results of ab initio molecular orbital calculations were used to determine the molecular mechanics constants for bond stretching motions, angular distortions and the important torsional motions. This approach has been successfully used by Army researchers to evaluate the toxicity of chemical agents and other related compounds.[5] This approach works well within the MM2 framework, but the increased complexity of the MM3 force field will likely require more data derived from ab initio sources. For instance, it is not clear that only force constant estimates will suffice to

derive an adequate MM3 model for novel compounds such as those encountered in high energy density materials.

An alternative to the MM force fields is provided by the Dreiding model.[6] This force field has a simpler mathematical form than does the MM force field and has recently been extended to include all atoms on the periodic chart, in principle.[7] In practice, many atoms are poorly described since there is little data available to permit a reasonable fitting. On the other hand, there are sufficient data to describe the atoms normally encountered in energetic materials applications. While the Dreiding model has been extensively used in the study of nucleic acids and biopolymers, we are not aware of applications to molecules containing the functional groups peculiar to energetic materials.

In addition to these molecular mechanics approaches, we have used both the semiempirical molecular orbital method, MOPAC,[8] and the ab initio electronic structure theory computer codes available on the BRL supercomputers to provide more data on the materials studied in this research. The MOPAC calculations were performed using the MMADS system of molecular modeling computer codes. This package also contains a modified version of the MM2 force field suitable for studies of molecules containing the phosphorus, sulfur, and chlorine atoms. These modifications are described in Reference 5. Most of the ab initio calculations have been performed using CADPAC.[9] We have described previous applications this software in previous Army Science Conference talks,[10] and in several recent papers.[11] The ab initio calculations provide accurate data on the molecular structure of model polymers, compounds containing two or three polymer units. The CADPAC codes also provide a substantial amount of information about the molecular properties of these model polymers. In the case of dielectric polymers, for instance, we are interested in those molecular properties that relate to the electronic charge distribution.

The following section describes the results of computations performed using the full spectrum of computational tools available in our laboratory. We note that the ab initio computations were performed using the Cray X/MP and Cray 2 computers at the Ballistic Research Laboratory. The MOPAC and MM2 calculations were performed using the MMADS software as installed on the Alliant FX/8 minisupercomputer in the Interior Ballistics Division computer facility. Molecular mechanics computations using the Dreiding force field, with various modifications as described, were done using the POLYGRAF software package installed on an SGI 4D240 installed in the Vulnerability and Lethality Division computer site. The POLYGRAF software was made available by Biodesign, Inc. for evaluation purposes.

DESCRIPTION OF THE COMPUTATIONS: To evaluate the suitability of molecular modeling techniques as predictive tools in energetic materials design, we have performed a range of calculations in two distinct applications areas. We first describe the results obtained in modeling the structure of the dielectric polymer poly(vinylidene fluoride). This material, and related compounds, are likely to be important in high energy density storage applications. Additionally, we have studied several compounds used as ingredients in propellant formulations.

The ab initio calculations employed atomic basis sets that have, in past applications, predicted structural parameters that agreed well with experimental data. Reference 11 provides several examples to demonstrate this point. As will be seen, the limitation to small model compounds does not inhibit the utility of these powerful methods.

POLYMER APPLICATIONS: Poly(vinylidene fluoride), henceforth referred to as PVDF, has become the most heavily studied piezoelectric polymer. PVDF $[-CH_2-CF_2-]$ and its copolymers exhibit large dielectric constants and piezoelectric and pyroelectric coefficients. These properties can, in principle, be computed using ab initio electronic structure methods. However, bulk properties and behavior are beyond the application range of such powerful theoretical tools. To describe a polymer in bulk we must use molecular mechanics methods. The question we wish to answer is, "Are the ab initio data and the molecular mechanics data consistent with one another?" One measure of such consistency is provided by the internal coordinates of the materials studied.

The structural parameters determined for two model compounds of PVDF, $C_4H_8F_4$ and $C_5H_8F_4$ are shown in Table 1. These computations illustrate the need to include enough of the polymer skeleton so that the end effects of the short chain do not dominate the structure. The C_5 compound is just large enough to show the constant bond lengths of the internal CC bonds. Both sets of calculations correctly predict the CCC bond angle dependence on the atoms bonded to the central carbon. Selected structural parameters predicted for $H-[CH_2-CF_2]_6-H$ using MOPAC and the Dreiding force field with various assumptions concerning the distribution of electrons in the molecule are shown in Table 2. These results correspond to the model for which all the C-C-C-C torsions were approximately 180 degrees. Examination of the results suggests that the overall agreement among the methods is quite good. All of the models predict the carbon-carbon bond that terminates the chain to be smaller than the internal CC bonds. The ab initio computations and all of the molecular models predict the internal CC bond lengths to be equal. Surprisingly, the MOPAC computations predict alternating CC bond lengths. This failure is critical and strongly hints that the method will not be useful in studies of these materials.

Table 1. Internal Coordinates for $C_4H_6F_4$ and $C_5H_8F_4$

	$C_4H_6F_4$	$C_5H_8F_4$
RC1 C2	1.503	1.504
RC2 C3	1.572	1.513
RC3 C4	1.505	1.513
RC4 C5	.....	1.504
RC1 H	1.079	1.079
RC2 F	1.387	1.387
RC3 H	1.080	1.080
RC4 F	1.379	1.387
RC4 H	1.073	.....
RC5 H	.....	1.079
C1 C2 C3	114.6	114.6
C2 C3 C4	117.3	117.7
C3 C4 C5	.....	114.3
C1 C2 C3 C4	180.0	180.0
C2 C3 C4 C5	.....	180.0

Table 2. Selected Internal Coordinates for $C_{12}H_{14}F_{12}$*

	MOPAC	Q-equil	DREIDING Del Re Q	Gasteiger
RC1 C2	1.585	1.535	1.545	1.546
RC2 C3	1.596	1.544	1.552	1.553
RC3 C4	1.601	1.541	1.551	1.554
RC4 RC5	1.595	1.544	1.553	1.554
RC5 C6	1.605	1.538	1.551	1.554
RC6 C7	1.599	1.538	1.551	1.554
RC7 C8	1.594	1.538	1.552	1.554
RC1 H	1.106	1.092	1.092	1.091
RC2 F	1.357	1.369	1.374	1.375
RC3 H	1.111	1.093	1.093	1.092
RC4 F	1.361	1.370	1.374	1.375
C1 C2 C3	109.8	115.7	117.6	117.8
C2 C3 C4	120.7	109.7	110.6	110.0
C3 C4 C5	113.1	116.3	117.7	118.5
C1 C2 C3 C4	180.0	180.0	180.0	180.0
C2 C3 C4 C5	180.0	180.0	180.0	180.0

*Units: Angstrom and degrees.

5

The three molecular modeling approaches differ little in the predicted parameters. All predict bond lengths that are longer than those predicted by the ab initio computation of the C_5 compound. Each model predicts a $CC(H_2)C$ bond angle that is close to the ab initio value, but all the $CC(F_2)C$ bond angles are too low. The molecular models also predict the CH bonds to be longer than those obtained in the ab initio calculations. This data suggests that the Dreiding force field predictions of molecular structures are primarily independent of the atomic charge model used. Electrostatic effects do not seem to affect the predicted structure in any significant way. However, the atomic charges, shown in Table 3, predicted by the equilibration of charge model[12] provides the best correlation with the atomic electron populations predicted by the ab initio calculations. Charges predicted by the Del Re[13] charge model agree with the alternation of charges predicted by the ab initio calculations, but the magnitudes are low. The charges predicted by the Gasteiger model are not in agreement with the ab initio results, especially since all the carbon atoms posess partial positive charges. Since the differences between the charge models will have significant effects upon efforts to model the bulk properties of polymers like PVDF, our results suggest that the equilibration of charge model will provide the most realistic electron distribution for computations done with the Dreiding force field.

Table 3. Predicted Atomic Charges for $C_4H_6F_4$

	Ab Initio	Q-Equil	Del Re	Gasteiger
C1	0.48	0.75	0.31	0.25
C2	-0.28	-0.10	-0.02	0.07
C3	0.78	0.78	0.36	0.26
H(C1)	0.07	0.13	0.07	0.10
F(C1)	-0.27	-0.45	-0.21	-0.21
H(C2)	0.13	0.67	0.65	0.04
F(C3)	-0.29	-0.42	-0.21	-0.21
C4	-0.42	-0.09	-0.02	0.08

The computational cost of determining optimized structures for sizable polymer fragments increases rapidly with complexity of the theoretical method. Although we would prefer to use ab initio methods, this is not possible when studying the interaction of polymers, nor if one is modeling the properties of copolymers. Given that only a molecular modeling approach is feasible, our results on the poly (vinylidene fluoride) infer that the Dreiding force field gives reasonable results when the either the equilibration of charge or Del Re charge model is employed. This is a hopeful result, since the Dreiding

6

model supports more atom types than does the MM3 model. Thus, we can hope that the model will work for numerous classes of chemical compounds, a conjecture that we now test with a particularly difficult organic species.

ENERGETIC MOLECULES: Diazidonitrazapentane (DANPE) is an energetic liquid used as a component in propellant formulations. To understand the difference in performance between DANPE and other compounds, we need to develop a relationship between the molecular structure of these compounds and their performance. Knowing the molecular structure is a prerequisite for determining structure-property relationships. Consequently, we have completed a full structural optimization for DANPE using our ab initio electronic structure codes. The full optimization required approximately one hundred hours of CPU time on the Cray X/MP-48 computer, including the computation of the second derivatives. Without the automatic potential energy searching algorithms available in most modern electronic structure theory codes, this optimization would have been impossible. Additionally, it is unlikely that any empirical investigation could have provided molecular structure parameters as accurate as those presented in Table 4. The listed parameters are the bond lengths and angles for one half of the heavy atom skeleton of DANPE, that is, the unique bond lengths and angles that characterize the heavy atom skeleton. The ab initio data is the only data on the structure of this unique compound. Given the computational expense of computing the structure, however, it should be clear that ab initio methods must be reserved for particularly crucial molecules. Here we consider whether the molecular modeling tools under consideration are capable of predicting structures for materials like DANPE which contain functional groups characterized by novel bonding situations. DANPE, of course, contains two azido groups. Azides have traditionally posed problems for electronic structure theories that do not consider the full electronic Hamiltonian operator.

In Table 4 we present the structural parameters predicted by the molecular models. On this case, we use the Gasteiger model to assign the partial atomic charges. The Del Re model cannot be used, in this case, since it has not been parametrized for some of the atom types needed to describe the azido groups. In addition, we have also used the partial atomic charges predicted by the ab initio calculations with the Dreiding model. The results presented in Table 4 clearly show that the empirical predictions are not accurate. While the coordinates for the nitramine portion of the molecule are predicted to be close to the ab initio results, neither force field calculation comes close to predicting the carbon-carbon bond length, or the carbon-azido nitrogen bond length. Furthermore the predictions for the two NN bonds in the azido group are exceptionally poor. Clearly, the empirical force field model employed does not contain parameters that adequately represent the

7

novel bonding that occurs in this compound. The most obvious source of error in obtaining these results is the atomic radius used to describe the sp-hybridized nitrogen atom. Clearly, accurate modeling of materials containing these novel subgroups will require extensions to the existing force fields. Such extensions can be developed using ab initio computations as the source of force field data. In fact, for novel functional groups, ab initio results will likely be the only reliable source for such information. This approach has been discussed by Hoffinger and others,[4] and has been employed by our colleagues at CRDEC.[15]

Table 4. Selected Internal Coordinates for DANPE*

	Ab Initio	Dreiding Gasteiger	Ab initio Q
RNN(O)	1.323	1.335	1.334
RNO	1.237	1.177	1.176
RCC	1.458	1.545	1.549
RCN(N)	1.481	1.397	1.399
RNN(N)	1.241	1.010	1.011
R(N)NN	1.109	1.003	1.002
ANNC	115.17	126.27	127.1
ANCC	113.41	110.43	112.5
ACCN	112.65	113.00	115.5
ANNO	117.65	123.52	123.5

*Units: Angstrom and degrees.

SUMMARY: We have presented detailed computational results that describe the structure and properties of several compounds important in modern high energy density materials research. Both state-of-the-art ab initio methods and modern molecular mechanics tools coupled to a powerful three-dimensional graphics display have been used. We have used the Dreiding force field model because the more familiar MM2 force field does not contain atom types that are relevant to the structure of the energetic materials studied. Except for the diazido compound, there was good agreement between the molecular modeling predictions and the ab initio data. Given that we could optimize the structure of a C_{12} fragment of PVDF in a few seconds, while ab initio computations of the structure of a C_5 fragment required ca. 10 hours of supercomputer time, it should be clear that the systematic study of a class of materials requires the application of the empirically based methods.

On the other hand, the ab initio tools available today are a unique source of molecular properties information. For instance, while the molecular models used to predict the structure of PVDF worked well, the ab initio calculations provided expectation values for many properties, including polarizability and susceptibility data. More sophisticated electronic structure codes, such as BROOKLYN89[16] can provide accurate information on such important processes as ionization and breakdown. Thus, in the effort to evaluate the performance of materials, one must have access to the full range of computational tools.

As a final comment we note that several other groups in this laboratory have had the opportunity to use the molecular modeling software that we used in our research. The methods have also been used to investigate interactions between materials presently employed in liquid propellant applications[17] and in low vulnerability ammunition (LOVA) applications.[18] Our experience has been that the tools are useful, but that the most important ingredient to the utility is the excellent graphics interface provided by the commercial package. Fortunately, we have been able to install the software on a Silicon Graphics 4D240 workstation equipped with state of the art graphics hardware. This powerful compute and graphics engine played an exceptionally important role in our studies.

REFERENCES:

1. N.L. Allinger, Y.H. Yuh and J-H. Lii, J. Am. Chem. Soc., Vol. 111, p. 8551, 1989.

2. J-H. Lii and N.L. Allimger, J. Am. Chem. Soc., Vol. 111, p. 8566, 1989.

3. J-H. Lii and N.L. Allinger, J. Am. Chem. Soc., Vol. 111, p. 8576, 1989.

4. A. Hofinger and R. Pearlstein, J. Comp. Chem., Vol. 5, p. 486, 1984.

5. J.M. Leonard and G.R. Famini, "A Users Guide to the Molecular Modeling, Analysis, and Display System", CRDEC-TR-030, January, 1990.

6. S.J. Weimer, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profets and P. Weiner, J. Am. Chem. Soc., Vol. 106, p. 765, 1984.

7. S.L. Mayo, B.D. Olafson, and W.A. Goddard, III, private communication.

8. J. Stewart, MOPAC, Version 3.1, United States Air Force Academy, 1983.

9. R.D. Amos and J.E. Rice, "CADPAC: The Cambridge Analytic Derivatives Paackage", issue 4.1, Cambridge, 1989.

10. G.F. Adams, R.T. Kroutil, N.C. Handy and R.D. Amos, "The A Priori Synthesis of Infrared Spectra for Chemical Agents", Army Science Conference, 1988.

11. G.F. Adams, M.M. Gallo and M. Page, Chem. Phys. Letters, Vol. 1162, p. 4497, 1989.

12. A.K. Rappe and W.A. Goddard, III, in press.

13. G. Del Re, J. Chem. Soc., Vol. 1958, p. 4031, 1958.

14. G. Del Re, B. Pullman, and T. Yonezawa, Biochem. Biophys. Acta, Vol. 75, p. 153, 1963.

15. G. Famini, private communication.

16. D.R. Yarkony, BROOKLYN89 Users Manual, 1989.

17. N. Klein, work in progress.

18. P.J. Kaste, work in progress.

Resin Modified Pavement (U)

Mr. Randy C. Ahlrich* and Mr. Gary L. Anderton
USAE Waterways Experiment Station
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

INTRODUCTION

A. Background

Asphalt concrete pavements are very susceptible to damage when subjected to fuel and oil spillage and severe abrasion from tracked vehicles. Over 80 percent of the Army's pavements are surfaced with asphalt concrete. Due to the mission of the Army and the equipment in its inventory, Army pavements are routinely subjected to fuel damage and severe abrasion. Tank trails, crossings, hardstands, wash facilities, motorpools, helicopter refueling pads, and aircraft parking aprons are examples of Army pavements that are susceptible to abrasion damage. A cost-effective surfacing material other than portland cement concrete is needed for construction and rehabilitation of Army pavements.

A resin modified pavement (RMP) was developed in France in the 1960's as a fuel and abrasion resistant surfacing material. The French company, Jean Lefebvre, developed this pavement process as a cost-effective alternative to portland cement concrete. The RMP process has been used on various types of pavements including warehouse floors, tank hardstands, and aircraft parking aprons. The resin modified pavement has been implemented in numerous countries including Great Britain, South Africa, Japan, Australia, and Saudi Arabia with good success.

The resin modified pavement is best described as a semi-rigid, semi-flexible pavement. The RMP is a tough and durable surfacing material that combines the flexible characteristics of an asphalt concrete material with the fuel, abrasion, and wear resistance of a portland cement concrete. The resin modified pavement process is basically an open-graded asphalt concrete mixture containing 25-30 percent voids which are filled with a resin modified cement slurry grout. The open-graded asphalt mixture functions as a support layer and determines the thickness of the RMP. The

slurry grout is composed of portland cement, fine aggregate, water, and a resin additive. The grout material is poured onto the open-graded asphalt mixture after the asphalt material has cooled, squeegeed over the surface, and vibrated into the voids with a small (3-5 ton) vibratory roller. The curing period can vary between 1 and 28 days depending on the type of portland cement used in the grout and the loading conditions. Figure 1 is a cut cross-section of a RMP field core.
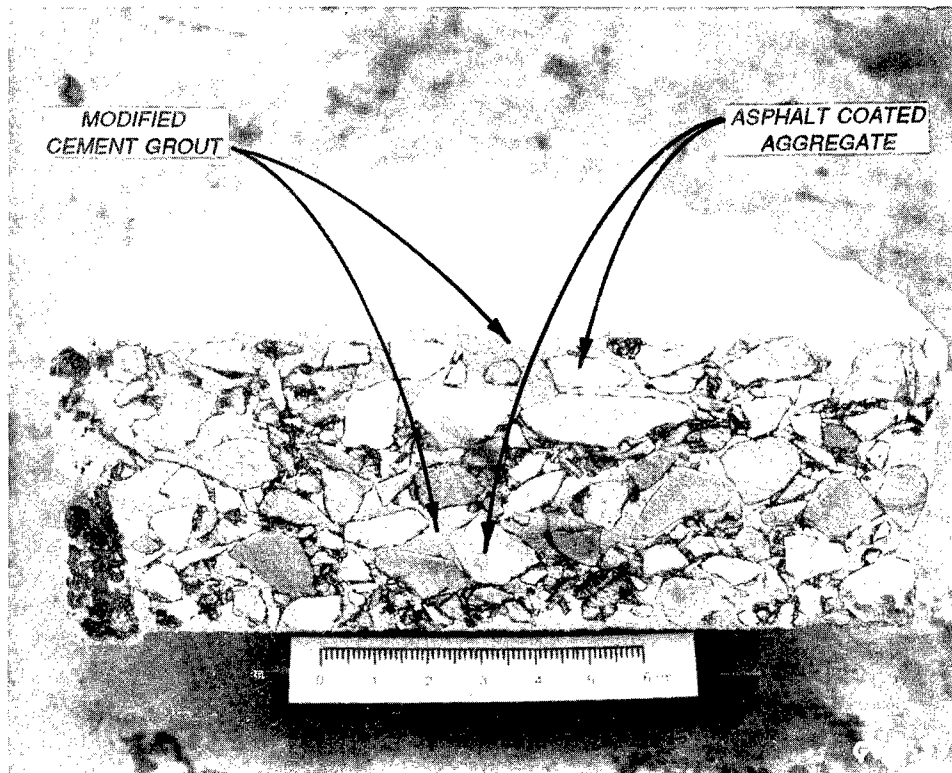
Figure 1.   Cross-section of RMP field core

During the mid 1970's, the U.S. Army Engineer Waterways Experiment Station (WES) evaluated the resin modified pavement. A test section was constructed to evaluate the effectiveness of this special surfacing material to resist damage caused by fuel, oil spillage, and abrasion from tracked vehicles. The results of this evaluation were not favorable. The test section did not resist damage caused by tracked vehicles and fuel spillage. The evaluation indicated that the effectiveness of the RMP was

very construction sensitive and if all phases were not performed correctly, the RMP process would not work.

In 1987, the Office, Chief of Engineers (OCE) tasked WES to reevaluate the resin modified pavement process. Good field performance in Europe and improved materials and construction procedures indicated this process had potential to be a viable alternative to standard paving materials. The evaluation began with an extensive literature search and background analysis into the RMP process. The review indicated that the majority of the in-service pavements constructed with this process were in Europe, especially in France. Site inspections were conducted to evaluate the field performance of several private and military RMP applications in France, Great Britain, and Australia. Visual observations of these sites indicated that the RMP process had tremendous potential for U.S. military applications.

B. Objective

The objective of this research was to determine the effectiveness of the resin modified pavement in resisting damage caused by severe abrasion from maneuvering tracked vehicles and from fuel and oil spillage. Recommendations on the potential future uses of the resin modified pavement would be made based on its determined effectiveness.

C. Scope

In order to determine the effectiveness of the resin modified pavement process, a RMP test strip was constructed at WES. A 150-ft. by 50-ft. test strip was constructed by a local contractor, APAC of Mississippi, with technical guidance from representatives of the parent company, Jean Lefebvre. The RMP test strip was constructed according to the specifications without any problems. The test strip was allowed to cure for 28 days to obtain full strength before any traffic was placed on the pavement.

The resin modified pavement was trafficked with the M-1 and M-60 tanks. Straight passes and 180 degree pivot steer turns were applied with the tracked vehicles to evaluate the abrasion resistance of the RMP. Five different fuels and oils (including jet aviation fuel, gasoline, diesel, synthetic oil, and hydraulic oil) were spilled on the RMP. Thirty cycles of controlled fuel and oil spillage were used to evaluate the fuel resistant properties. The Federal Highway Administration's Accelerated Loading Facility (ALF) was also used to traffic the RMP. ALF simulated heavily-loaded, high tire pressure truck traffic. Data drawn from the traffic tests and fuel resistance analysis would provide the basis for the recommendations made for this new pavement process.

PRE-CONSTRUCTION ANALYSIS

A. Site Evaluation and Thickness Design

An old pavement testing area near the Geotechnical Laboratory at WES was chosen as the construction site for the proposed 150-ft. by 50-ft. resin modified pavement test strip. Various types of asphalt concrete test pavements had been constructed in this area by WES researchers since World War II. In 1982, approximately 1 1/2-in. of asphalt concrete was placed over the entire area as a leveling course. Since that time, the area had been used occasionally to stockpile aggregates and soil samples.

Since the resin modified pavement is essentially a surface course, there was a need to ensure that the underlying layers of the test strip would be structurally sound. If any load related failures were to occur during future traffic tests, a structurally sound foundation would leave no doubt that the failure was initiated in the surface course, which was the pavement layer being evaluated in this case. A high quality bituminous mixture was determined to be the most economical means of obtaining the structurally sound foundation needed.

A thickness design procedure was conducted to determine the required thickness of this asphalt concrete layer. The first step of the design procedure was to determine the strength properties of the existing pavement. Nondestructive tests (NDT) were conducted using the falling weight deflectometer (FWD) for this purpose. The pavement deflection data captured during these tests were input to a computerized layered-elastic program (BISDEF) which computes strength properties and predicts elastic moduli for each pavement layer. The moduli values were then input to a computerized pavement thickness design program (AIRPAVE) to determine strengthening overlay requirements. A design load near equivalent to the M-1 and M-60 tank loads was used in this thickness design program. The results of this exercise indicated that a total thickness of 3 in. of asphalt concrete surface mixture would provide a structurally sound foundation for the 2-in. thick resin modified pavement wearing course.

B. Materials Evaluation

1. Dense-Graded Bituminous Intermediate Course

As previously mentioned, a high quality bituminous surface mixture was selected as the best means of providing a sound foundation beneath the resin modified pavement surface course. The aggregate gradation specified for this pavement layer was that which is listed in the U.S. Army Standard Practice Manual (TM 5-822-8) for a 3/4-in. maximum size, high tire pressure

16

surface blend. An AC-30 grade of asphalt cement was specified for the dense-graded intermediate course.

Once the contract for constructing the test strip was awarded, the contractor was asked to provide samples of the aggregates and asphalt cement to WES so that the materials could be tested and a job mix formula developed. The job mix formula presented to the contractor before construction began contained the specific aggregate gradation and asphalt cement content desired. These and other requirements for the dense-graded intermediate course were specified in the final construction specifications. The section of the final specifications relating to the intermediate course was designed after the standard U.S. Army Guide Specification for Bituminous Intermediate and Surface Courses for Airfields, Heliports, and Tank Roads (CEGS-02556).

2. Open-Graded Asphalt Mixture

A review of the available literature indicated that the mix design of the open-graded asphalt mixture would play a critical role in the proper construction of the resin-modified pavement. The majority of the mix design was found to focus on the final voids content of the compacted open-graded asphalt mixture. The general requirement is 25-30 percent voids in the final compacted mixture. Any amount less than this would not allow the slurry grout to fully penetrate the open-graded mixture, resulting in a structurally unsound surface course which would likely have excessive cracking and deterioration. Void contents greater than this amount would increase the cost of the pavement without providing any significant structural improvements and could also reduce the pavement strength by eliminating some of the aggregate to aggregate interlock.

A laboratory analysis of the open-graded asphalt mixture was conducted prior to construction of the test strip to determine the amount of asphalt cement that would produce the proper void content in the final mixture. Although checks against such aggregate properties as fractured faces and particle shapes had to be made, the final void content of the compacted mixture was the main focus of this laboratory analysis. The majority of the laboratory mix design guidance found in the literature was based on French methods which use non-traditional specimen sizes and compaction methods. Therefore, a preliminary study was first conducted in the WES laboratories to determine the best mix design method in terms of standard U.S. military practices. This study indicated that the standard Marshall mix design procedure, which is both a military standard and an ASTM standard with 20 blows of the hand hammer for the compactive effort would indicate the proper asphalt content to achieve the required voids criteria. A subsequent mix design study followed involving a wide range of asphalt contents. The asphalt cement used in the laboratory study and actual

construction was the same type as that used in the dense-graded intermediate course, an AC-30 grade. The aggregate gradation used was taken from the literature as the standard gradation for heavy-duty pavement applications. This gradation is listed below.

Open-Graded Asphalt Mixture Aggregate Gradation

US Standard Sieve Size	3/4"	1/2"	3/8"	#4	#8	#30	#200
% Passing	100	67	44	22	12	5	2

The results of the open-graded mixture laboratory analysis indicated that an asphalt cement content of 4.0 percent would result in a void content of near 30 percent in the final compacted mixture. This asphalt content, along with the aggregate gradation shown above, was specified in the final construction specifications as the open-graded asphalt mixture job mix formula.

3. Resin Modified Slurry Grout

A pre-construction laboratory study was also performed on the resin modified slurry grout. The available literature was fairly specific about the types of materials and relative proportions of these materials to produce a satisfactory slurry grout. Nonetheless, laboratory tests were deemed necessary to ensure that these recommendations would work for the materials which were to be used in the test strip.

The individual components of the slurry grout are cement, sand, filler, water, and a cross polymer resin of styrene and butadiene. The type of cement used is purely a design option, as is the case for portland cement concrete. WES used a standard Type I cement. The sand must be clean, sound, durable, and range in size from the No. 30 to No. 200 sieve sizes. WES used a washed silica sand to meet these requirements. The filler must have a very fine gradation (minimum of 95 percent passing the No. 200 sieve) and may be fly ash, limestone dust, or rock flour. WES used a fly ash. The resin additive acts as a plasticizer to reduce the slurry grout viscosity for better penetration and as a strength producing agent. The solid constituents of the grout are near equal proportions of sand and filler with about twice that amount of cement. Enough water is added to produce a water to cement ratio of 0.60 to 0.70. The resin additive is added to the mixture last in an amount equal to 2.0 to 3.0 percent of the total batch weight. This combination of ingredients produces a slurry grout which is very fluid and only slightly more viscous than water.

The laboratory analysis conducted at WES on the slurry grout consisted of varying the mix proportions within the recommended allowances to determine the best mix formula. The single acceptance criteria for the slurry grout is a Marsh cone viscosity of 7.0 to 9.0 sec. immediately after mixing. (Water has a Marsh cone viscosity of 6.0 sec.) Because this viscosity range is relatively narrow, slight variations of the water to cement ratio and amount of resin additive were used to obtain a slurry grout mix formula of the proper viscosity.

After 10 different slurry grout formulations were mixed and tested in the laboratory, a final formula was derived which was found to produce a slurry grout viscosity of just over 7 sec. It was thought that a slurry grout in the lower end of the acceptable viscosity range combined with an open-graded support layer in the upper end of the acceptable voids range would help to ensure full penetration of the grout during construction. The final slurry grout formulation used on the test strip is listed below.

Resin Modified Slurry Grout Formula (by weight)

Type I Cement	38.2 %
Fly Ash	19.1 %
Sand	13.3 %
Water	26.7 %
Cross Polymer Resin	2.7 %

CONSTRUCTION

A. Dense-Graded Bituminous Intermediate Course

The construction of the test strip began with the dense-graded bituminous intermediate course. The existing surface was swept clean and a light tack coat of Type SS-1 asphalt emulsion was sprayed on the clean surface by a distributor truck. The tack coat was used to bond the new dense-graded asphalt mixture with the existing asphalt surface. This tack coat was applied during the afternoon before construction of the intermediate course began to allow for enough curing time and to prevent construction delays the next morning.

With the construction equipment already in place, the intermediate course construction was completed in less than 1 day. The hot mix was spread with a mechanical paver and compacted with a 10-ton rubber-tired roller and a 10-ton steel-wheeled roller. Samples of the hot mix were taken at several intervals during the day for determination of mixture properties by WES laboratory personnel. These laboratory quality control

tests, along with data obtained from field cores cut out of the test strip early the next day, indicated that both the mix and construction procedures were satisfactory. A final thickness of approximately 3 in. was laid across a 160-ft. by 60-ft. area. These dimensions were designed to provide the sound foundation required for the 2-in. thick, 150-ft. by 50-ft. resin modified surface course.

B. Resin Modified Pavement

1. Open-Graded Asphalt Mixture

The open-graded asphalt mixture was placed on top of the dense-graded bituminous intermediate course the day after the intermediate course was placed. A light tack coat was sprayed onto the intermediate course using the same type of asphalt emulsion and application rate as before. The tack coat was allowed to cure for a few hours before the open-graded mixture construction began.

Similar to the quality control techniques used during the construction of the intermediate course, samples of the hot open-graded mix were taken from the haul trucks at several intervals during the day. Laboratory tests on these materials to determine the asphalt content, aggregate gradation, and most importantly the final void content, helped to determine the properties of the in-place mix. Additionally, core samples were cut out of the hardened test strip the following morning to again check these same properties. All loose mix samples and core samples indicated that the open-graded mix was placed with satisfactory material properties and construction techniques.

The open-graded mixture was spread with the same mechanical paver that was used for placing the intermediate course. Under normal circumstances, open-graded mixes of this nature tend to cool off relatively quickly because of their high internal voids. This means that the required compaction usually must follow closely behind the paver which is placing the mix. Because the ambient temperatures were so unusually high during the construction of the test strip, rapid heat loss of the asphalt mix was not a problem. To the contrary, the afternoon temperatures which reached well over 100 degrees F forced the construction crews to wait several hours before compacting so that the roller would not cut and shove the hot mat.

As is the case for most open-graded asphalt mixes, compaction during construction was not used to achieve any density requirements, but merely to "seat" the asphalt coated aggregates and smooth over the rough surface. A relatively small (3-ton) steel-wheeled roller was used to accomplish this. The static, light-weight steel-wheeled roller is used as opposed to

the more traditional heavier models (8-ton to 10-ton) so that a minimal loss in voids is incurred during the compaction process.

Once the open-graded asphalt mixture had cooled for several hours, a single pass of the small steel-wheeled roller in the static mode was made over the entire 150-ft. by 50-ft. area. Small cut marks were left along the edge of the roller wheels after this process. Therefore, after another hour of cooling, another pass of the small roller was used to roll out these marks. After these final passes of the roller, the construction of the open-graded asphalt layer was complete. The entire freshly paved area was covered with polyethylene sheeting for the night to prevent contaminants (dirt, sand, etc.) from blowing onto the pavement surface and falling into the open voids.

2. Resin Modified Slurry Grout

The resin modified slurry grout was added to the open-graded asphalt pavement the day after the open-graded mix was placed. The slurry grout used in the construction of the test strip was made at a local cement batch plant. The dry cement, sand, and fly ash were mixed in the plant's pugmill (in their proper proportions) for several minutes before dumping into the transient mixer truck. Next, the proper amount of water was dumped into the transient mixer truck and the resulting slurry grout was mixed in the rotating mixing drum for several minutes. At this point, the proper amount of the cross polymer resin additive was poured into the mixing drum and the truck operator was allowed to leave the plant site and head toward the test strip job site while the mixing truck continuously rotated in-transit.

Once at the test strip job site, the transient mixing trucks were allowed to position themselves directly on the open-graded asphalt pavement which had hardened overnight. A sample of the slurry grout was first taken and the Marsh cone viscosity was checked on the test strip job site to ensure that the grout was of the proper viscosity. Samples were taken from each transient mixer truck at the test strip job site and approved before the grout was placed.

The grout was to be placed in the same 10-ft. wide longitudinal lanes as were used during construction of the open-graded asphalt pavement. This pattern would give a sense of order to the grout application and prevent over-working of the hand working crew. This crew consisted of four people working broom-handled squeegees behind the transient mixer truck. The slurry grout was slowly poured onto the open-graded asphalt surface and when an area became saturated with grout, the squeegees were used to pull the grout along the surface to undersaturated areas. The grout was poured onto the pavement surface after traveling down a pivoting delivery chute. As the grout was being slowly dumped onto the pavement, one person

continuously directed the chute to dry areas of the pavement. Once an area of a lane was completely saturated with grout, the truck driver slowly moved the truck forward. After a short time at the beginning of the grout application, the squeegee operators, chute operator, and truck driver were able to continue the grouting procedure in an efficient, controlled manner. Figure 2 is a typical view of the grout application procedure used on the test strip.

Figure 2. Resin modified slurry grout being applied to
open-graded asphalt mixture

Immediately behind the grouting operation, the small steel-wheeled roller made several passes over the grout filled pavement in the vibratory mode in order to ensure that all subsurface voids were being filled. Because the void content of the open-graded asphalt pavement and the slurry grout viscosity were within the specified ranges, the vast majority of the internal voids were filled with grout simply by saturating the pavement when the grout was poured over the surface. There did seem to be, however, a small amount of internal voids isolated from the initial grout

application as evidenced by small air bubbles which appeared behind the vibratory roller as it passed. These air bubbles usually only appeared after the first pass of the vibratory roller, indicating that all voids were being filled with grout.

After each of the five 10-ft. lanes had been saturated with grout and vibrated, all excess grout remaining on the surface was removed by continually pulling the hand squeegees in one direction. This process also served to fill any possible undersaturated areas. After this final step, the grout application was complete.

3. Curing

After the grout application was completed, a curing compound was sprayed over the surface of the wet, grout-filled pavement. The material used was a white pigmented concrete curing compound which is commonly used in curing Type I portland cement concrete. The white pigments are used to reduce maximum pavement temperatures during the curing period. This in turn reduces the expansion and contraction stresses resulting from extreme temperature changes. An overabundance of these stresses can lead to shrinkage cracking during the curing period. The grout was applied by a pressurized, hand operated sprayer wand with a fan type nozzle. A light coating of the curing compound (200 sq. ft. per gal.) over the entire test strip completed the construction process. The pavement was allowed to cure with no traffic applied for 28 days before traffic testing began.

EVALUATION

A. Tracked Vehicle Traffic

As previously mentioned, the resin modified pavement test strip was allowed to cure for 28 days before any traffic was allowed on the pavement. This cure time was allowed to ensure that the RMP had plenty of time for adequate strength gain. The effectiveness of the RMP greatly depended on its performance during the tracked vehicle trafficking.

Tracked vehicle traffic on the RMP test strip consisted of M-1 (gross weight 113,000 lbs.) and M-60 (gross weight 100,000 lbs.) tanks. Six hundred 180 degree pivot steer turns at the same point and 5000 straight passes were applied with the tracked vehicles to the test strip. Excessive track wear was noticed during the initial trafficking of the RMP. During the initial stage of trafficking, the RMP withstood the abrasive action of the pivot steer turns very well; only excess grout was worn off. As the tank track turned during the pivot steer, the track pads would drag across the RMP surface (Figure 3). After 420 turns, the tracked vehicle produced enough rough abrasion and high stresses to start surface ravelling. The

surface ravelling began without any warning. Once the ravelling started, the deterioration increased rapidly because the loose debris that had been dislodged was now being dragged and scraped across the RMP surface causing further damage.
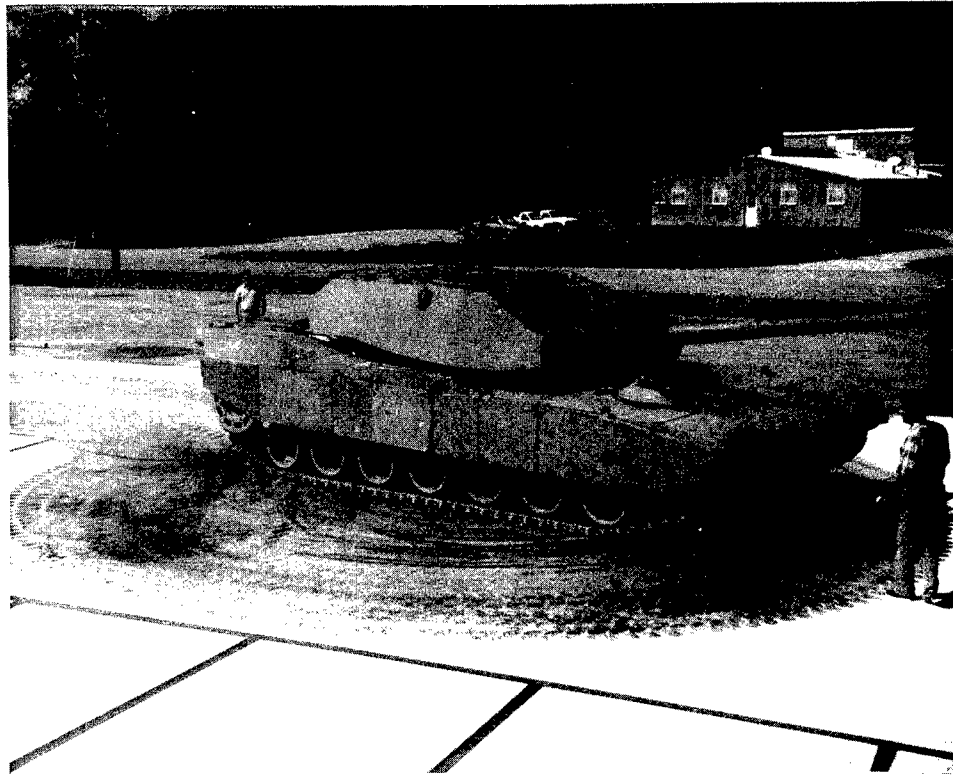
Figure 3. M-1 tank performing pivot steer turns on RMP test strip

At 600 pivot steer turns, the turning traffic was stopped because the abrasive action had produced a ravelled area 1-in. deep covering 35-sq. ft. It was thought that concentrated pivot turns of this nature are not commonly found in the field, making this traffic test much more severe than normal applications. As an example, it would require the tanks of four armored divisions performing 180 degree pivot steer turns at the same exact point to equate to this traffic test.

The 5000 straight passes with the tank traffic only caused slight surface wearing of the grout, which exposed the surface of the coarse aggregate. The tracked vehicle moving forward and in reverse caused no significant damage to the RMP. At the conclusion of the tracked vehicle

trafficking, it was determined that the resin modified pavement had effectively demonstrated a resistance to severe traffic abrasion and could be used as a pavement surface for tracked vehicles.

B. Fuel and Oil Spillage

Five different fuels and oils were used to evaluate the effectiveness of the RMP in resisting deterioration caused by fuel and oil spillage. Jet aviation fuel, gasoline, diesel, synthetic oil, and hydraulic oil were spilled on the resin modified pavement. Thirty cycles consisting of 1-qt. of each material were spilled on the RMP surface. Each spill area was 8-ft. by 10-ft. The materials were spilled from a height of 30 in. The rate of spillage was set so that each material took 20-30 min. to drip 1 qt. The fuels and oils were allowed to set on the RMP for an additional 30 days. Figure 4 shows the condition of the fuel and oil spillage areas after this 30-day period.

G609- 218

Figure 4. Fuel and oil spillage areas

25

Visual observations indicated that the RMP was resisting deterioration from fuel and oil spillage. However, field cores taken from the spillage areas indicated that the fuels and oils had penetrated the resin modified pavement. Slight deterioration was observed in the oil spills. The gasoline and jet aviation fuels had a fast rate of evaporation which prevented these materials from significantly penetrating the RMP. The diesel fuel penetrated the RMP the fastest and caused the most damage. Once again, this test is thought to be an acceleration of typical fuel spillage problems in the field as most spills are normally cleaned and not allowed to soak into the pavement for several months.

After the fuel and oil penetration had been discovered, the stability of the resin modified pavement was questioned. The maximum penetration was approximately 1-in. in the diesel area. The remaining fuels and oils penetrated less than 1/2 in. A 1-ton van was used to traffic the fuel spillage areas. Fifty passes and fixed position, power steering turns were applied to the contaminated areas by the van. Only slight scuffing was noticed after the van had trafficked the RMP with no appreciable damage.

C. Accelerated Loading Facility (ALF)

The Federal Highway Administration's Accelerated Loading Facility (ALF) was also used to traffic the resin modified pavement test strip (Figure 5). ALF simulated truck traffic by applying a load of 19,000 lbs. to a dual wheel assembly with tire pressures of 140 lbs. per sq. in. ALF applied 80,000 passes to a 48-in. wide strip of the RMP. No appreciable deterioration or deformation occurred in the wheel path. Only slight wearing of the excess grout on the RMP surface was observed. The ALF evaluation indicated that vehicular traffic had little effect on the resin modified pavement and that the RMP should have excellent field performance when trafficked by rubber-tired vehicles.

SUMMARY

The resin modified pavement construction process can be used to build new pavements or rehabilitate existing pavements that are subject to heavy, abrasive loads and fuel spillage. The RMP can be used to surface areas used by tracked vehicles such as tank trails and crossings, hardstands, and wash facilities. This pavement may also be used in motorpools, refueling pads, and aircraft parking aprons. The RMP provides an alternative surfacing material in areas where conventional pavement materials have excessive maintenance problems. The resin modified pavement can be used in place of asphalt concrete and portland cement concrete in these specialized areas.

Figure 5. The Accelerated Loading Facility trafficking the RMP test strip

The RMP provides a tough and durable surfacing material for military pavements. The current data and evaluations indicate that the resin modified pavement process has potential for several pavements uses. The initial cost of the RMP is estimated to be between $10 and $15 per sq. yd. as compared to $15 to $25 per sq. yd. for portland cement concrete with equally traded vehicle abrasion resistance. At this price, the RMP is a cost-effective method to construct or rehabilitate many of the Army's abrasion and fuel-resistant pavements.

Chemistry of Solid Propellant Ignition

*William R. Anderson, Dr.
Anthony J. Kotlar, Dr.
George F. Adams, Dr.
Ballistic Research Laboratory, SLCBR-IB-I
Aberdeen Proving Ground, MD   21005-5066

I. INTRODUCTION: The study of the detailed chemistry which takes place in premixed laminar flames of intermediates formed in the combustion of solid gun propellants has been an area of intense interest to the energetic materials community for the last 10 years. Such study yields information on chemical mechanisms directly applicable to models for steady-state combustion of the solid propellants. The rate of reaction of the gas phase intermediates just above the burning propellant surface affects the temperature gradient above the surface and, thus, the all-important rate of heat feedback to the condensed phase propellants. This rate in turn controls the overall burn rate (linear regression rate) of a propellant strand. Detailed chemical mechanistic study of the steady-state combustion of solid propellants has recently been the focus of research by Hatch[1] and Melius.[2] The degree of success of these models, which contain rather gross assumptions about the difficult to probe interface region between solid and gas phases, has served notice of the usefulness of this type of approach and stirred much interest in the energetic materials community. Comparatively less attention has been given to the related problem of the modeling of chemistry relevant to ignition of solid propellants. Because of ignition delay problems noted in certain situations with low vulnerability (LOVA) propellants (in particular XM39) in large caliber guns, there is currently much interest in this type of time dependent problem.[3] It is thought that the problem will cause aiming difficulties and could even lead to breechblows. Of particular interest, we note that ignition frequently takes place in the gas phase intermediates, apparently formed by pyrolysis of the propellant, at some point far from the propellant surface. Detailed modeling to identify key reactions and species which control the chemical induction period prior to full ignition in the gas phase would be extremely useful. We envision using the results of a simple, homogeneous model of the chemistry taking place during the induction period to eventually yield reduced mechanisms which could be used in

multidimensional models of gun ignition, i.e., the XNOVAK code.[4] In general, the chemical reactions to which the results are sensitive are expected to be different for the two types of problem (homogeneous ignition vs steady-state combustion). Therefore, although these problems are closely interrelated, both types of study are needed.

This paper will focus primarily on our preliminary efforts to model ignition of a simple mixture, H_2/NO_2. We plan in the future to see whether the detailed chemical results for chemical induction (ignition) delay time can be sensibly modeled with a reduced reaction set. Preliminary work seems to indicate a favorable outlook for success. The results for H_2/NO_2 include parametric studies on the several mixtures, mainly centering on the effects of initial temperature on the results. Of significant interest is the result that large concentrations of highly reactive radical species (which in general control the overall rate of combustion) present in the initial mixture have little effect on the predicted ignition delay times. Preliminary calculations on CH_2O/NO_2 mixtures, which more closely approximate expected pyrolysis products of propellants, exhibit the same insensitivity to radical concentration. This result suggests it may be possible to ignore radicals leaving the surface of a pyrolyzing propellant in future models of large caliber gun ignition. This is important because the initial levels of these radicals cannot be measured easily.

II. GAS PHASE IGNITION MODEL: All of the calculations were performed utilizing the CHEMKIN package of subroutines from Sandia Laboratories, Livermore, CA.[5] This highly versatile package of subroutines relating to chemical kinetic modeling of large networks of reactions may be used in a large variety of problems. A very simple model was used for the ignition of gas phase propellant intermediates. (The model was adapted from a sample problem given at the end of the CHEMKIN manual.[5]) We assume a zero dimensional, homogeneous distribution of all reactants. In addition, we assume constant pressure, adiabatic conditions. Given an initial condition (i.e., set of concentration of reactants, temperature and pressure) and a chemical mechanism, including well-established rate constants, it is then possible to solve for the time dependent temperature, species concentrations, and heat release rate. In addition, one can determine the molar rate of production (or destruction) of a given species by each reaction at any point in time. From this information, the chemical pathways of the overall mechanism for a given set of conditions may be readily determined. The equations solved are:

$$\frac{dT}{dt} = -\frac{1}{\rho \bar{c}_p} \sum_{k=1}^{K} h_k \dot{\omega}_k W_k \tag{1}$$

30

$$\frac{dY_k}{dt} = \frac{\dot{\omega}_k W_k}{\rho} \quad k = 1,2,\ldots,K \tag{2}$$

where T is temperature, t is time, ρ is density, c_p is constant pressure specific heat, h_k is specific enthalpy of species k, W_k is the molecular weight of species k, and Y_k is the mass fraction of species k. (For those not used to thinking in terms of specific enthalpies, note that $h_k W_k$ is simply the molar heat of formation of species k at temperature T.) Note that there are K species, resulting in K+1 differential equations. These are solved using the integration method discussed in the sample problem.[5]

III. RESULTS AND DISCUSSION: As previously stated, most of our discussion will center on the H_2/NO_2 system. As a prototypical example, much of the effort was spent examining results for starting conditions of $T_0 = 1000K$, an equivalence ratio (fuel/oxidant) of 1.00 (stoichiometric), and a pressure of 1 atm. In our first report on this work,[6] the bulk of the mechanism was obtained by abstracting all of the reactions (and their rate constants) involving the species H, N, and O from a model for NO_x formation in air-supported combustion of hydrocarbons.[7] A number of reactions were then added to the mechanism to account for the fact that our problems involve a regime of much higher concentration of NO_x species. More recently, we have revised this mechanism. One may arrive at the mechanism used to obtain the present results by again starting with the H, N, O reactions from Ref. 7. The following reactions are added to the mechanism thus obtained:

R1. $2 NO_2 = 2 NO + O_2$

R2. $2 NO_2 = NO + NO_3$

R3. $2 NO_3 = 2 NO_2 + O_2$

R4. $2 NO_2 + M = N_2O_4 + M$

R5. $NO_2 + O + M = NO_3 + M$

R6. $NO_3 + NO_2 = NO + NO_2 + O_2$

R7. $NO + HO_2 = HNO + O_2$

R8. $H_2 + NO_2 = HONO + H$

R9. $OH + NO + M = HONO + M$

R10. $OH + NO_2 + M = HNO_3 + M$

R11. $NO + M = N + O + M$

R12. $NO + H_2 = HNO + H$

R13. $HNO + O = OH + NO$

31

R14. $HNO + NO_2 = HONO + NO$

R15. $2 HNO = H_2O + N_2O$

R16. $HONO + O = OH + NO_2$

R17. $HONO + OH = H_2O + NO_2$
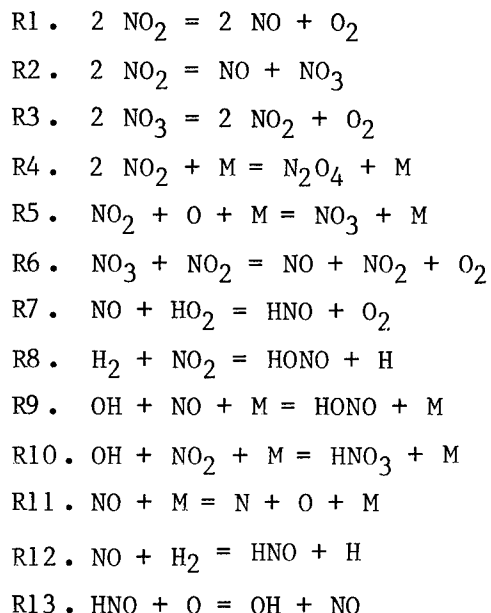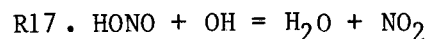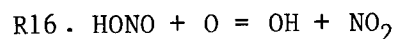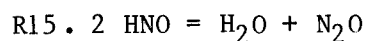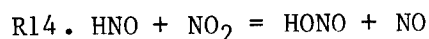
In addition, the following reactions (or their reverses) were included in Ref. 7. Their rate constants have been revised as discussed below.

R18. $NO + O = N + O_2$

R19. $NO_2 + M = NO + O + M$

R20. $H + NO + M = HNO + M$

R21. $H + NO = N + OH$

R22. $OH + HNO = H_2O + NO$

R23. $N_2O + M = N_2 + O + M$

R24. $N_2O + O = N_2 + O_2$

R25. $N_2O + O = 2 NO$

R26. $H + N_2O = N_2 + OH$

R27. $H + NO_2 = NO + OH$
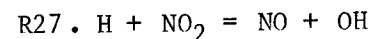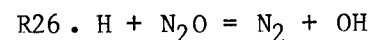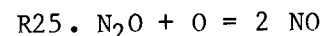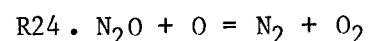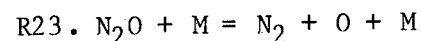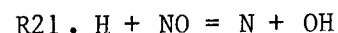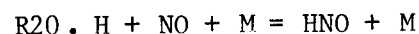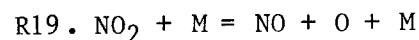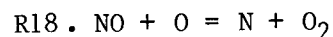
Rate constants for reactions R4 and R6 were obtained from the Leeds review.[8] The rate constant for R3 was estimated from data discussed in the Leeds review as $k = 2.6 \times 10^{12} \exp(-3875/T)$. (Fortunately, this reaction is not very important because little NO_3 forms.) The rate constant for R7 was taken from an estimate by Howard, as discussed by Hanson and Salimian.[9] The rate constant for R8 was taken from Ref. 10. Rate constants for the rest of the reactions listed above were taken from the recent review of Tsang and Herron.[11] For all of the reactions involving a collider, M, the zero pressure limiting rate constant was used and the collider was assumed to be N_2. The resulting mechanism consists of 76 reactions involving 23 species. It is worth noting that the rate of reaction R8 was actually determined in the only time dependent ignition study known to us on this system, the shock tube study of Slack and Grillo described in Ref. 10. The full mechanism described here does a reasonable job of predicting the ignition delay times observed for the mixtures used in the shock tube study.

Concentration and temperature vs. time curves for our prototypical example for H_2/NO_2 are shown in Figs. 1-3. One notes immediately that there is a distinct two stage nature to this time dependent combustion, which is typical of NO_2 oxidized combustion. For instance, on examination of the temperature curve (Fig. 1), one notes the temperature

stays at a fairly constant, low value for a while until it suddenly begins to rise sharply at about 80 microsec and climb to an intermediate temperature of about 2370 K. The temperature curve then abruptly changes slope and much more slowly climbs towards the adiabatic flame temperature of about 3100 K. Figure 2, the majority species plots, shows that the sharp upward portion (first stage) of the temperature curve corresponds to the consumption of NO_2 and half of the H_2. Almost all of the NO_2 is rapidly converted to NO during this time period and, for reasons which will be discussed momentarily, almost none of the NO reacts until the NO_2 is consumed. The sharp break in the temperature curve corresponds to the point in time when virtually all of the NO_2 is gone. It is worth noting that a calculation of the adiabatic temperature reached by this mixture excluding N_2 as a final product (i.e., an "equilibrium" mixture in which all of the nitrogen is forced to remain in the form of NO) is within 50 K of the temperature at the break point. At this point, the NO begins to react slowly (second stage) until the final products are obtained. Several trace species profiles are shown in Fig. 3. These show that OH and H are important at all stages after ignition begins, but that O does not become as important until the second stage is reached. Prior to ignition, O is not as important either, although this is not obvious because of the scale of the plot. It is also not obvious that HO_2 is very important prior to ignition. In fact, conversion of the radical pool (total of H, O, and OH concentrations - O is not very important during the preignition stage, as will be discussed later) to HO_2 has an important inhibiting effect on the ignition. Finally, also shown in Fig. 1 is the heat release rate, i.e., the evolution of heat due to chemical reaction, vs. time. Note that the heat release rises to a sharp maximum during the first stage of combustion. For purposes of discussion, we have defined the term "chemical induction time" or "ignition delay time" as the time at which this maximum is observed. (Note that if the term ignition delay time is used in reference to this model, it has a quite different meaning from the same term used in reference to a gun's ballistic cycle.)

In spite of the simplicity of our homogeneous combustion model, much can be learned about the combustion chemistry of the mixture under study which can be applied to steady-state flame conditions. This results because the reactions important once ignition occurs (that is, at and after the steep temperature gradient) are usually the same set as are important in steady-state flames. Also, a study of H_2/NO_2, in particular, is of considerable interest because its reaction mechanism is expected to be a subset of the reaction mechanisms necessary to describe the ignition and combustion of many solid propellants. This is the case because NO_2 is produced as an important intermediate in the pyrolysis or combustion of these propellants. Examination of the species formation and production rates through the individual reactions

33

leads to Fig. 4. Figure 4 depicts the reactions responsible for the conversion of reactants to products (according to the model) for our prototypical example. During the first stage of combustion, NO_2 is converted to NO almost exclusively by the reaction $H + NO_2 \rightarrow NO + OH$. Meanwhile, half of the H_2 is converted to H_2O by the reaction $OH + H_2 \rightarrow H_2O + H$. During the second stage, the NO is converted to N_2 via several reactions of comparable importance. Almost none of the NO formed during the first stage reacts during the first stage. It cannot be overemphasized how strong this effect is. The reason for the effect is that the reduction of NO to N_2 at the rate predicted in the second stage requires a fairly high concentration of H atoms (in addition to other radicals whose concentrations are linked to that of the H atom) because the rate constants for its various reactions shown in the figure are all rather small. Now, the reaction of H with NO_2 is one of the fastest neutral species combustion reactions known. It is so fast that the concentration of H atoms is held very low when even a small fraction of NO_2 remains in the mixture. After the NO_2 has dropped to a value close to its equilibrium concentration (much too low to be seen in Fig. 2), the H atom concentration suddenly begins to rise very sharply (because the fast $H + NO_2$ reaction is no longer possible) and the NO begins reacting via $H + NO \rightarrow N + OH$ (and other competing reactions). Depending on initial conditions, one may see a rise in the H atom concentration of 3-6 orders of magnitude at this point in time. The chemical reactions important during the time leading up to ignition will be discussed later.

A parametric study of the chemical induction (ignition) delay time is given in Table 1. Shown are the resulting delays (τ) for a variety of initial temperatures, and a few interesting pressures, equivalence ratios ϕ, and assumed initial starting mole fractions of H atom, X_H. Upon inspection of the $\phi = 1.0$, P = 1 atm portion of the table with zero initial H atom present, one notes immediately that there is a very strong dependence of the delay time on the initial temperature of the mixture. It in fact changes almost 8 orders of magnitude between initial temperature, T_0, of 500 and 1500 K. This is not surprising when one notes that the main initiation reaction, R8 as will be discussed in a moment, has an activation energy (Arhenius E_a factor) of about 29.0 kcal/mole.[10] This is large enough to cause a considerable temperature sensitivity. Upon changing to $\phi = 1.50$ (the fuel/oxidant equivalence ratio of the neat cyclic nitramines RDX and HMX, typical LOVA propellant ingredients), the delay is observed to drop slightly (at least for $T_0 = 1000$ K). It is possible that this is due to the lower heat capacity of a mixture containing more H_2 vs. NO_2, which could lead to a larger temperature rise during the early stages of the reaction for an almost equivalent amount of heat release. Upon further changing to a pressure of 10 atm, approximately equivalent to the pressure observed during the early ignition stages of large caliber guns,[12] the delay

34

ANDERSON, KOTLAR & ADAMS

observed for an initial temperature of 1000 K is observed to drop about
a factor of 6 from its value at 1 atm. This surely results because of
the increase in the rate of the main ignition reaction, R8, at the
higher pressure due to its bimolecular nature. The effects of added H
atom will be discussed later.

Table 1. Chemical Induction (Ignition) Delay for H_2 NO_2 Mixtures
(The precision of these calculated delays is better than 1%)

T_0 (K)	$X_H = 0$	$X_H = 10^{-4}$	$X_H = 10^{-3}$	$X_H = 10^{-2}$
$\phi = 1.0$, P = atm				
500	158 s	150 s	95.5 s	2.15 s
560	9.00 s			
600	1.84 s			
700	66.5 ms			
800	3.94 ms			
900	382 μs			
1000	81.5 μs	78.0 μs	53.0 μs	7.8 μs
1100	30.6 μs			
1200	14.5 μs			
1300	7.85 μs			
1400	4.70 μs			
1500	3.02 μs	3.02 μs	2.74 μs	1.35 μs
$\phi = 1.5$, P = 1 atm				
1000	70.0 μs	65.5 μs	39.0 μs	4.26 μs
$\phi = 1.5$, P = 10 atm				
1000	12.8 μs	12.3 μs	9.05 μs	0.675 μs

The homogeneous ignition model we are using at present does not
have any capabilities for sensitivity analysis. Thus, it is difficult
to determine which elementary reactions have the most effect on the
predicted ignition delay times. However, it was possible by using
chemical intuition to select a number of reactions which one might
suspect to be important because of their nature without a laborious
procedure, such as changing the rate constants one by one, say by a
factor of 10, then repeating the ignition delay calculation to see which
reactions have a large effect. (It is one of our primary goals to bring
a sophisticated sensitivity analysis to bear on this type of problem.)
One notes that in this type of overall reaction, in which one starts
with no radicals present in the so-called radical pool (i.e., the total

of H, O, and OH concentrations, the radicals which carry the fast chain reactions and control the overall rate at which the reaction proceeds) elementary reactions which add or remove radicals from the radical pool are expected to be very important. For the assumed mechanism, there are only three reactions which can possibly produce radicals at $t = 0$:

R8. $H_2 + NO_2 = HONO + H$,

R28. $H_2 + M \rightarrow H + H + M$,

R29. $NO_2 + M \rightarrow NO + O + M$
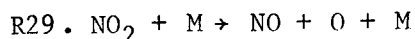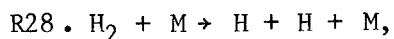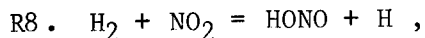
Examination of the rates of reactions which can introduce or remove radicals from the radical pool (not merely change their identity) just after $t = 0$, but before full ignition, revealed the following reactions could also be important:

R30. $OH + NO_2 \rightarrow HO_2 + NO$

R31. $HO_2 + OH \rightarrow OH + OH$

R32. $HO_2 + OH \rightarrow H_2O + O_2$

R10. $OH + NO_2 + M \rightarrow HNO_3 + M$

R7. $NO + HO_2 \rightarrow HNO + O_2$

R33. $HONO + M \rightarrow OH + NO + M$
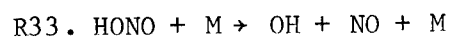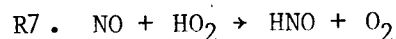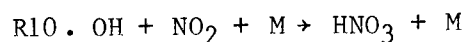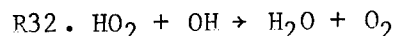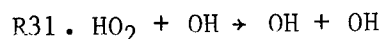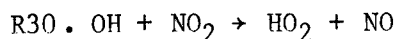
(Note that R33 is the reverse of R9. The calculation has revealed that R9 proceeds in the reverse direction during preignition stages.) R8, R28, and R29, if important, would lead directly or rapidly via subsequent reactions to the formation of H and OH, thus leading to ignition; R33 may happen after R8, enhancing ignition; R30 and R10 convert a highly reactive radical (OH) to a less reactive species, HO_2 or HNO_3, respectively, thus inhibiting ignition; finally, R31, R32 and R7 appear to determine whether the HO_2 formed in R30 is subsequently returned to the radical pool. Examination of the individual rates of R8, R28, and R29 reveals that R8 is the overwhelmingly predominant initial radical formation step at all times during the preignition stage. This is not at all surprising when one considers that the bond dissociation energies of H_2 and NO_2 are much larger than the activation energy for R8. In fact, removal of R8 from the mechanism causes the predicted delay time to increase by at least a factor of 1000, while the alternative removal of R28 or R29 has almost no effect at all. Removal of R33 from the mechanism results in an increase in the ignition delay time of about 10%, thus confirming its enhancement effect. [This removal is actually performed by removing R9 from the reaction mechanism. Removal of a reaction results in both the reaction and its reverse being ignored because the reverse reaction rate constants are

calculated using the forward rate constants and equilibrium constants for the reaction.] Removal of R10 results in a decrease of about 15% in ignition delay, thus confirming its inhibiting effect. Removal of R30, on the other hand, results in a very large effect; the ignition delay is reduced by about a factor of 2. This confirms that conversion of radicals, at least temporarily, to HO_2 has a strong inhibiting effect. The HO_2 may be returned to the radical pool by R31, or it may be removed by the termination step R32 or the termination sequence R7 followed by $HNO + OH \rightarrow H_2O + NO$ (examination of the rates affecting HNO reveal the latter reaction is the fate of most HNO). Surprisingly, removal of R31 had little effect on the ignition delay, in spite of the fact that this reaction has the fastest rate of those destroying HO_2. This apparent lack of sensitivity is not well understood. Removal of R32, on the other hand, results in a decrease in ignition delay of about 5%, confirming a slight inhibiting effect. Removal of R7 produced a very slight decrease in ignition delay. Increasing R7 by a factor of 10 results in an increase in ignition delay of about 10%. Thus, the reaction is on the threshold of producing an important inhibiting effect. It is important to note this, because the rate constant for this reaction is not well known. In particular, its activation energy was estimated from similar reactions to be about 2 kcal/mole; the high temperature rate constants were obtained by extrapolating from room temperature results using this E_a. If the activation energy could be as high as 5 kcal/mole, this would produce an increase in the extrapolated rate constant of about a factor of 10 at 1000 K. The reaction should be considered in ignition of other mixtures containing NO_2. Therefore, careful measurements or calculations of its rate constant at high temperature would be useful.

We also considered the effects of adding radicals to the initial mixture in this ignition problem. As previously stated, the highly reactive radicals H, O, and OH sustain the reaction once ignition occurs. It does not matter whether we assume highly reactive radicals are added as one or the other of these species in our initial mixture because they will rapidly interconvert amongst themselves via the fast reactions $H + NO_2 \rightarrow NO + OH$, $H_2 + OH \rightarrow H_2O + H$, and $H_2 + O \rightarrow OH + H$. Therefore we chose to examine the effects of added reactive radical only in the form of H atom. Let us return to the remaining columns in Table 1. We note immediately that a mole fraction of H atom of 10^{-2} yields a very pronounced decrease in the ignition delay for all starting conditions. This result is not surprising since this high a concentration is close to the radical concentration observed in most flames at temperatures above about 2000 K. However, a mole fraction of 10^{-3} yields a decrease of less than a factor of two, while a mole fraction of 10^{-4} has little effect on the predicted delays. For sake of comparison, consider a stoichiometric mixture of H_2/O_2 at P = 1 atm, T_0 = 1000 K, using exactly the same mechanism. For this case initial mole

fractions of 0.0 and 10^{-4} lead to predicted delay times of 0.133 and 0.027 millisec, respectively. Obviously, H_2/O_2 mixtures are much more sensitive to initial radical concentration. Examination of the H_2/NO_2 system with added H atom shows that most of the added reactive radical is rapidly being converted to HO_2, which is much less reactive with the majority species. This occurs because radicals in the H, OH, O pool are drained out through conversion of OH to HO_2 via R30. Perhaps the H_2/O_2 system is more sensitive to addition of radicals because the only comparable reactions which can occur to remove them in this system are three body reactions such as $H + H + M \rightarrow H_2 + M$, $H + OH + M \rightarrow H_2O + M$ and $H + O_2 + M \rightarrow HO_2 + M$. Such reactions are generally much slower than bimolecular reactions such as R30. In addition, the H_2/O_2 system has available two chain branching steps (two radicals produced for one consumed): $H + O_2 \rightarrow OH + O$ and $O + H_2 \rightarrow OH + H$. These could tend to make the system more sensitive to the addition of reactive radicals. Preliminary work on CH_2O/NO_2 mixtures, using a simple mechanism obtained from a recent shock tube study,[13] indicates a similar degree of insensitivity to initial reactive radical concentration. Now, a mole fraction of 10^{-4} for radical concentrations is much higher than one would predict for the equilibrium concentration of radicals in these mixtures at, say 1000 K. It is, in fact, much higher than the degree of dissociation of NO_2 at 1000 K. The gases evolved from a pyrolyzing propellant would be expected to be at a much lower temperature than 1000 K. Although the evolution of these gases is a dynamic process (as opposed to equilibrium) our work suggests that the radical concentrations in propellant gases in the case of long gun ignition delays (a problem with some low vulnerability propellants) are probably low enough that they can be assumed to have a negligible effect on calculated ignition properties. This is a very important and useful assumption because these cannot be measured or calculated easily, if at all, at present.

Lastly, shown in Fig. 5 are results of our preliminary efforts to reproduce results from our complex reaction network calculations using a reduced reaction set. Of course, the ultimate in reduction of the reaction set is to use one global reaction. We attempted this for our prototypical example using the simple functional form:

$$q = Q \, (m \rho)^{\nu} \, A \, \exp \, (-E/RT)$$

where Q is the total mass specific heat release of reacted starting materials, ν is the effective reactant order, $A \exp (-E/RT)$ is an Arhenius expression for the global rate constant with preexponential factor A and activation energy E, and m is the mass fraction of reactant remaining, given by:

$$m = m_0 \, [1-(T-T_0)/(T_f-T_0)]$$

where T_0 and T_f are initial and final temperatures, respectively and m_0 is the initial reactant mass fraction, taken to be 1. Note that this latter expression forces all the reactant to be consumed to final products when T reaches T_f. The fitted parameters are ν, AQ and E. Since we are mainly interested, ultimately, in predicting the ignition delay accurately with a simple functional form, and not the subsequent reaction history, we only attempted a fit of the first stage. (The reason for this is that once the ignition is started, the subsequent chemistry under conditions of high pressure, as in a gun, is expected to be so fast that it may be assumed to take zero time in the interior ballistic codes. It is clear that a reduced reaction set capable of reproducing a two stage structure would require at least two reactions involving an initial reactant, an intermediate, and a product.) The results in Fig. 5 are for our prototypical example which had T_0 = 1000 K and the stoichiometric ratio of H_2 and NO_2. In the figure, the heat release rate is replotted as a function of the temperature, rather than time. This was done because the functional form chosen is one of temperature rather than time. The fitted result for the first stage, shown in Fig. 5, is a close reproduction of the result from the detailed kinetic calculation. The fitted parameters are: ν = 0.729, AQ = 1.78 x 10^{10} (Q in cal/g and A in appropriate cgs units) and E = 22.6 kcal/mole. It is interesting to note that the fitted activation energy, E, for the global reaction is rather close to the activation energy, 29.0 kcal/ mole, of the most important radical producing ignition reaction, H_2 + $NO_2 \rightarrow$ HONO + H. We have not yet attempted to calculate ignition delays using the global reaction kinetics. It is, of course, of great interest to determine whether a single reaction with functional form for the rate constant as shown can reproduce the delays given in Table 1. Calculations using the H_2/NO_2 mechanism will thus provide a useful test of techniques to produce reduced reaction sets, prior to calculations for mixtures more closely resembling gases evolving from pyrolyzing propellants.

REFERENCES

1. (a) R.L. Hatch, "Chemical Kinetics Combustion Model of the
   NG/Binder System," Proceedings, 23rd JANNAF Combustion Meeting,
   CPIA Publication 457, Vol. I, p. 157, October 1986.
   (b) R.L. Hatch, "Chemical Kinetics Modeling of HMX Combustion,"
   Proceedings, 24th JANNAF Combustion Meeting, CPIA Publication 476,
   Vol. I, p. 383, October 1987.

2. C.F. Melius, "The Gas-Phase Flame Chemistry of Nitramine
   Combustion," Proceedings, 25th JANNAF Combustion Meeting, CPIA
   Publication 498, Vol. II, p. 155, October 1988.

3. G.E. Keller, "Report on JANNAF Workshop Influence of Gas-Phase
   Kinetics on Low Pressure Ignition and Flamespreading in Solid
   Propellant," BRL Technical Report BRL-TR-2918, July 1988.

4. P.S. Gough, "Modeling Chemical Interactions in Ignition of Gun
   Propellant," Proceedings of 22nd JANNAF Combustion Meeting, CPIA
   Publication 432, Vol. I, pp. 223-237, October 1985; see also
   G.E. Keller and A.W. Horst, "The Two Phase Flow Simulation of LOVA
   Propellant Interior Ballistic Behavior Using the XNOVAK Code," BRL-
   TR-2796, April 1987, and references therein.

5. R.J. Kee, J.A. Miller, and T.H. Jefferson, "CHEMKIN: A General-
   Purpose, Problem-Independent, Transportable, Fortran Chemical
   Kinetics Code Package," Sandia National Laboratories Report No.
   SAND80-8003, March 1980.

6. W.R. Anderson, A.J. Kotlar, and G.F. Adams, "Detailed Chemical
   Kinetic Modeling Relevant to Propellant Combustion," in
   Proceedings, 26th JANNAF Combustion Meeting, in press.

7. P. Glarborg, J.A. Miller, and R.J. Kee, "Kinetic Modeling and
   Sensitivity Analysis of Nitrogen Oxide Formation in Well-Stirred
   Reactors," Combustion and Flame, Vol. 65, p. 177, 1986.

8. D.L. Baulch, D.D. Drysdale, and D.G. Horne, Evaluated Kinetic Data
   for High Temperature Reactions. V.2. Homogeneous Reactions of the
   H_2-N_2-O_2 System, Butterworths, London, 1973.

9. R.K. Hanson and S. Salimian, "Survey of Rate Constants in the N/H/O
   System," in Combustion Chemistry, W.C. Gardiner, ed., Springer-
   Verlag, New York, 1985.

10. M.W. Slack and A.R. Grillo, "Rate Coefficients for H_2 + NO_2 = HNO_2 + H Derived from Shock Tube Investigations of H_2 - O_2 - NO_2 Ignition," Combustion and Flame, Vol. 31, p. 275, 1978.

11. W. Tsang and J.T. Herron, "Chemical Kinetic Data Base for Propellant Combustion: I. Reactions Involving NO, NO_2, HNO, HNO_2, HCN, and N_2O," submitted to J. Phys. Chem. Ref. Data.

12. G.E. Keller, BRL, private communication.

13. C.Y. Lin, H.T. Wang, M.C. Lin, and C.F. Melius, "A Shock Tube Study of the CH_2O + NO_2 Reaction at High Temperatures," submitted to Int. J. Chem. Kinet.
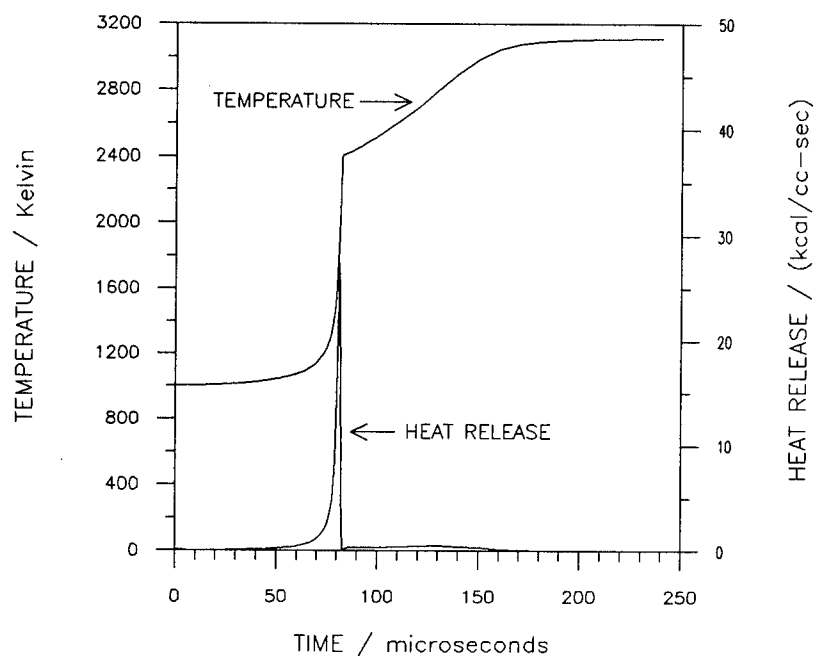
Fig. 1. Temperature and Heat Release Rate Vs. Time for an Adiabatic, Constant Pressure H_2/NO_2 Mixture with Initial Conditions $\phi=1.0$, P=1 atm
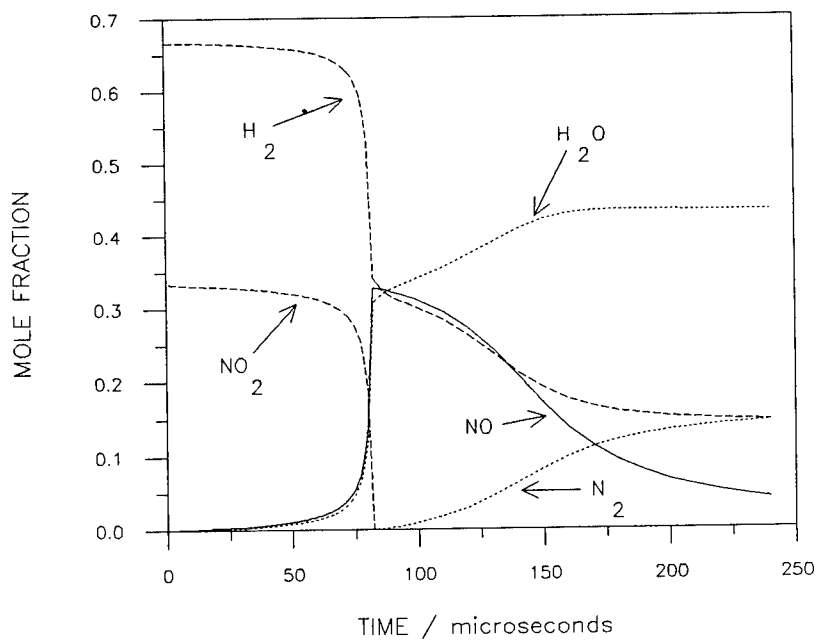
Fig. 2. Majority Species Vs. Time for an H_2/NO_2 Mixture with Conditions as in Fig. 1
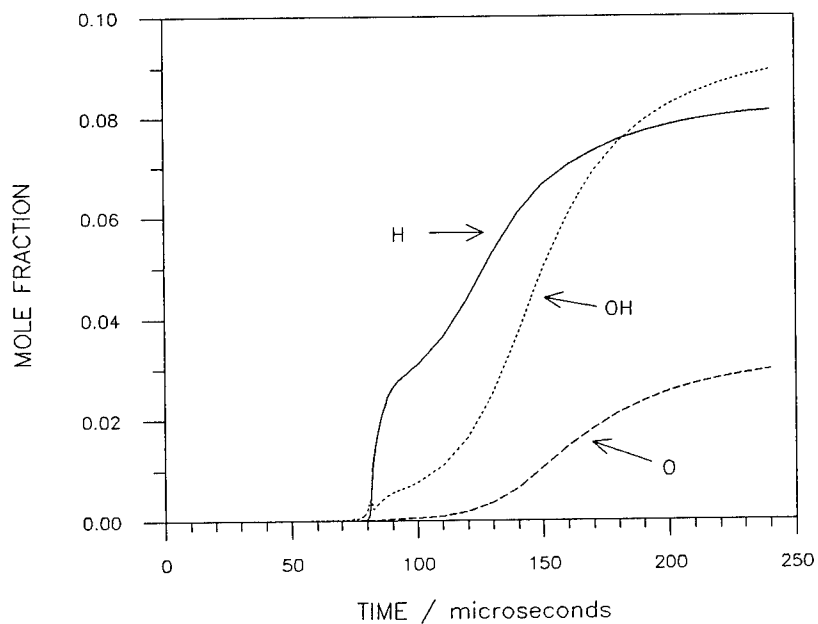
Fig. 3. H, O, OH Concentrations Vs. Time for an H_2/NO_2 Mixture with Conditions as in Fig. 1

H_2/NO_2 MECHANISM
Nitrogen Chemistry

STAGE 1   STAGE 2
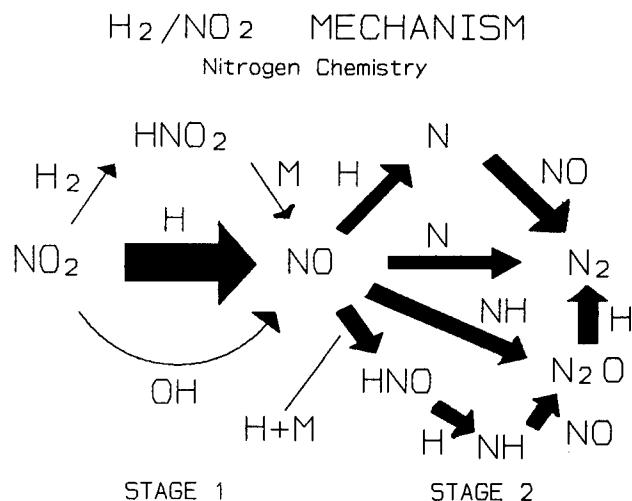
Fig. 4. Chemical Pathways for Majority Nitrogen Species in an H_2NO_2 Homogeneous Ignition Problem. The chemistry is that after full ignition has been reached (steep portions of the temperature-time curve). Large arrows indicate major pathways.
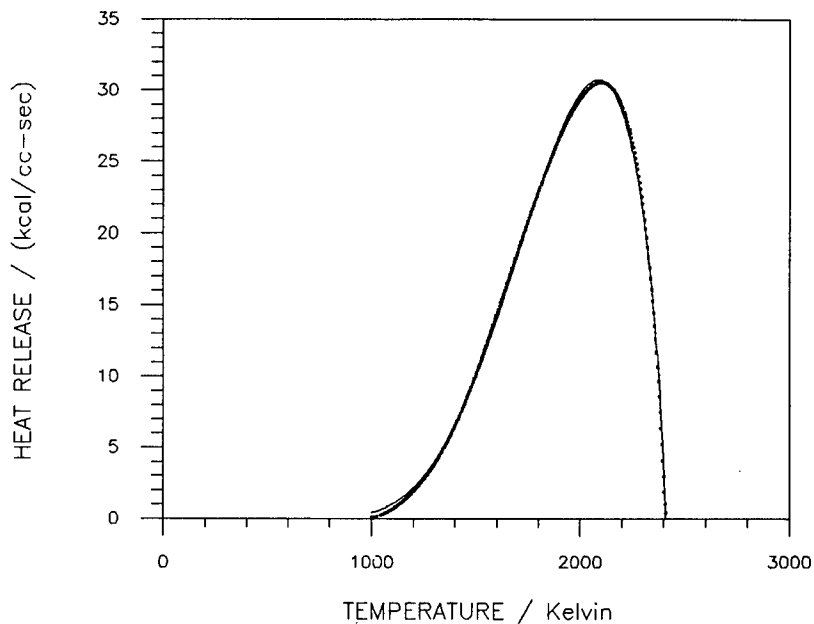
Fig. 5. Fit of the Single Global Reaction Model to the Results of a Detailed Reaction Network Calculation. (*) are from the detailed calculation, the solid line is the least-squares fit; conditions are those of Fig. 1.

43

Development of Vehicle Woodland Camouflage Patterns
Analogous to Background Contrast Highlights (U)

*George Anitole and Ronald L. Johnson
U.S. Army Belvoir Research, Development and Engineering Center
Fort Belvoir, Virginia 22060-5606

Christopher J. Neubert
U.S. Army Materiel Command
Alexandria, Virginia 22333-0001

1.0 Background

All military equipments have characteristic shapes, shadows, and highlights which contrast with the surrounding background and make the items conspicuous. The purpose of pattern painting is to disrupt these detectable signature cues using lusterless paint to reduce highlights, color to reduce contrast with the background and distort shadow areas, and patterns to distort the geometric shapes. Pattern painting reduces the threshold of detection and recognition, and provides a base for further camouflage.

The designs of current camouflage patterns have been developed subjectively based upon the designer's artistic determination of the necessary shapes, sizes, and colors to disrupt the signature cues. The basic guidelines in developing these patterns are to avoid designs which show up as vertical, horizontal, circular, parallel, or plaid type patterns. The object is to develop irregular shapes of similar contrast and size to those of the background as determined from selected ranges. There is no relationship between the colors and pattern shapes and the natural shapes, shadows, and colors found in a typical woodland background. Their determination is not based on any technical, scientific, or analytical process. As a result, the most effective patterns and colors for blending into a woodland background are not used. The development of such patterns has been a continuing empirical effort and applies not only to the U.S. but to foreign countries as well.

This paper describes a technical method of developing "natural" camouflage patterns which are analogous to the contrast highlights found

in woodland backgrounds. This includes the shape, size and color requirements. The pattern development process is described, along with the vehicle field test to determine the comparative effectiveness of the new pattern against four other patterns and a monotone vehicle.

2.0 Technical Approach

The procedure for developing the natural pattern is based on a photographic process of adjusting the contrast of background photographs to the point where only the major contrast highlight features are recognizable, and then adapting these features into pattern shapes for military equipment. The basic photographs used to develop this process consisted of treelines at distances chosen to enable acquiring sufficient working detail of the tree background. The photographs included both deciduous and coniferous treelines. Helicopter photographs were taken from oblique and vertical positions at different altitudes. The original photographs were taken on black and white high-contrast film and printed on high-contrast paper. The prints were re-photographed, and the process repeated until only the bold contrast image highlights remained after the minor contrast highlights and fine-tone details dropped out. The original photographs were converted into silhouette-like patterns by the interim steps of having progressively increased the contrast between light and dark areas, so that all the areas darker than a certain density were converted to black, and all the areas lighter than this density were converted to white. The intermediate gray tones dropped out. The bold contrast image highlights are representative of a given degree of light reflectively analogous to the bold contrast highlights found in natural foliated backgrounds.

Each final image silhouette was evaluated by magnification and compared to the other final images to determine commonalities between image silhouettes taken of various areas and tree types. The contrast image silhouettes showing the most correspondence were selected for further analysis. Final comparisons for commonalities resulted in the selection of four contrast image silhouettes for use in developing patterns for each side of the test vehicle. A portion of the higher contrast photographs, representative of an average blend of light and dark areas, was then selected as the basic pattern for each side of the vehicle, as shown in Figure 1. The circled area indicates the specific part of the image used in the pattern. The pattern for the top of the vehicle was developed in the same manner, using the aerial photographs. The selected black and white areas formed the basic silhouette shapes to be used in the camouflage pattern. Three density levels were chosen, including an intermediate level made up of small areas of black, but interpreted as gray shades on the high-contrast print. In developing the image silhouettes, it was found that very little difference existed

46

Area selected for pattern development for a given vehicle side
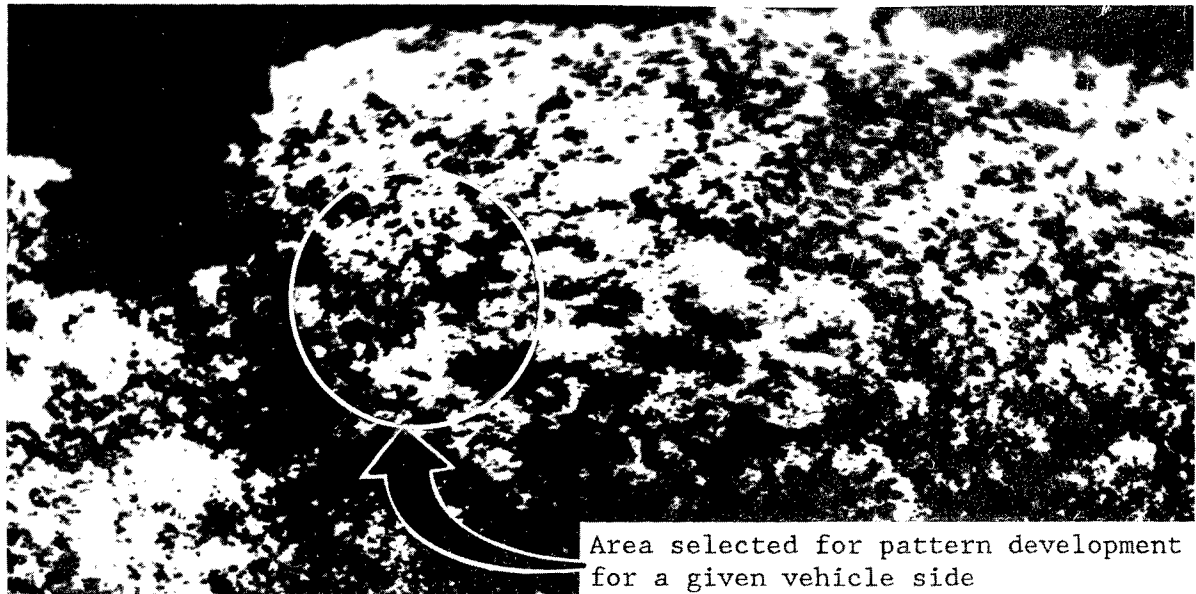
Figure 1. Sample Higher Contrast Photograph

between coniferous and deciduous treeline photographs, with the exception of slightly more lengthy horizontal high-contrast areas for coniferous backgrounds.

The size of the pattern shapes was determined by the condition that sufficient brightness contrast exists between the major colors at ranges between where an uncamouflaged object should become discernible from the background as the observer gets closer, to a close-in point where the shape and form of an object become obvious regardless of the pattern.

Color photographs and aerial video were used to determine the color scheme for the pattern black and white areas. Three color areas were evident. The two primary colors of light green and black were roughly equally divided, with the minor color of forest green taking about 15% of the area. This color combination appeared to best match the backgrounds investigated in color and brightness. The color combinations presented sufficient contrast to prevent merging the colors when viewed from long ranges and preserved the disruptive look at close ranges.

Past U.S. pattern designs failed to take into account the dark areas in a woodland background. When viewed close up, these areas are shades of gray or gray-brown. When viewed from a distance, these areas tend to be seen as black or dark areas. The treeline highlights are represented by

bright green, particularly in the sunlight, and were well depicted by the selection of light green. Black depicted the dark areas.

A scale outline drawing was made of the vehicle or equipment to be camouflaged, and the outlines of the selected black and white pattern portions were then suitably enlarged to fit the drawings based on previous pattern size determination. The pattern areas were hand-sketched on the full-scale equipment, which was then painted using lusterless paints in the 1% reflectance range. The pattern did not have to be applied with any strict degree of accuracy in any direction, because the contrast highlights found in nature are variable. Errors, or improvisation, were tolerated as long as the general design was followed.

3.0 Test Procedure

3.1 Test Patterns

The Natural Pattern was compared against four other patterns and a monotone. The patterns were applied to M-113 armored personnel carriers (APCs), and are shown in Figure 2. The four patterns and one monotone tested against the Natural Pattern were as follows:

A. U.S. Four-Color Pattern. The colors used on this pattern were the spring-summer woodland colors. The two major color areas were forest green and light green, and the two minor areas were black and field drab. The major colors each comprised about 45% of the pattern areas, and the minor colors about 5% each. The paint was lusterless, reflecting in the 1% range.

B. Dual-Tex Pattern. This pattern (developed by West Point) was derived from the U.S. Four-Color pattern, and followed the same general shapes and colors. The color patterns for the U.S. Four-Color pattern have been randomly disrupted to produce a higher textured appearance. The disruptive edges were created using a four-inch-square grid in a zig-zag manner.

C. Swedish Pattern. The Swedish pattern was characterized by the use of four lusterless (1%) colors with straight-line boundaries between the colors. The paints were formulated to match the Swedish specifications in color and reflectance. The pattern was 20% black with accents of light brown, and the rest of the surface equally divided between light green and dark green. The comparatively large percentage of black, as compared to the U.S. Four-Color pattern, created large contrasting areas when viewed from long distances. The straight-line shapes tended to soften with distance.

D. British Pattern. The British pattern was relatively simple, consisting of gray-green and black. Large bold paths of color were used on this pattern, with no particular guidance on shape, other than the pattern design should be carried up over the vehicle and pulled over the

U.S FOUR—COLOR PATTERN

BLACK
FOREST GREEN
LIGHT GREEN
FIELD DRAB

SWEDISH PATTERN

BLACK
LIGHT GREEN
LIGHT BROWN
DARK GREEN
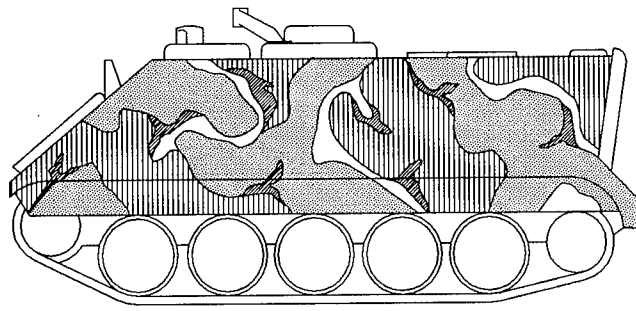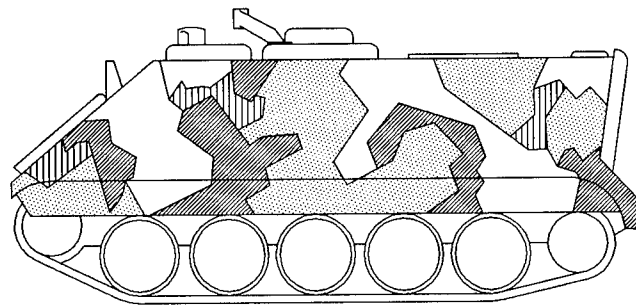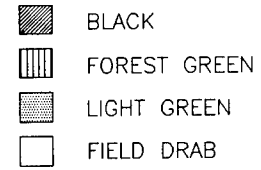
NATURAL PATTERN
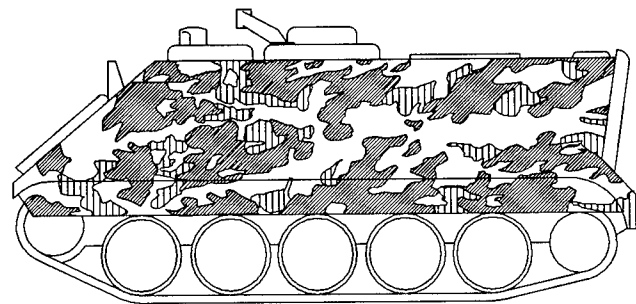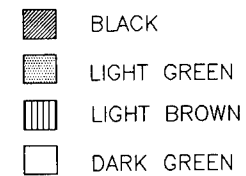
BLACK
FOREST GREEN
LIGHT GREEN
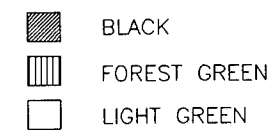
Figure 2. Candidate Test Patterns/Color

49

DUAL-TEX PATTERN

	BLACK
	FOREST GREEN
	LIGHT GREEN
	FIELD DRAB

BRITISH PATTERN

| | BLACK |
| | GREY-GREEN |

FOREST GREEN MONOTONE
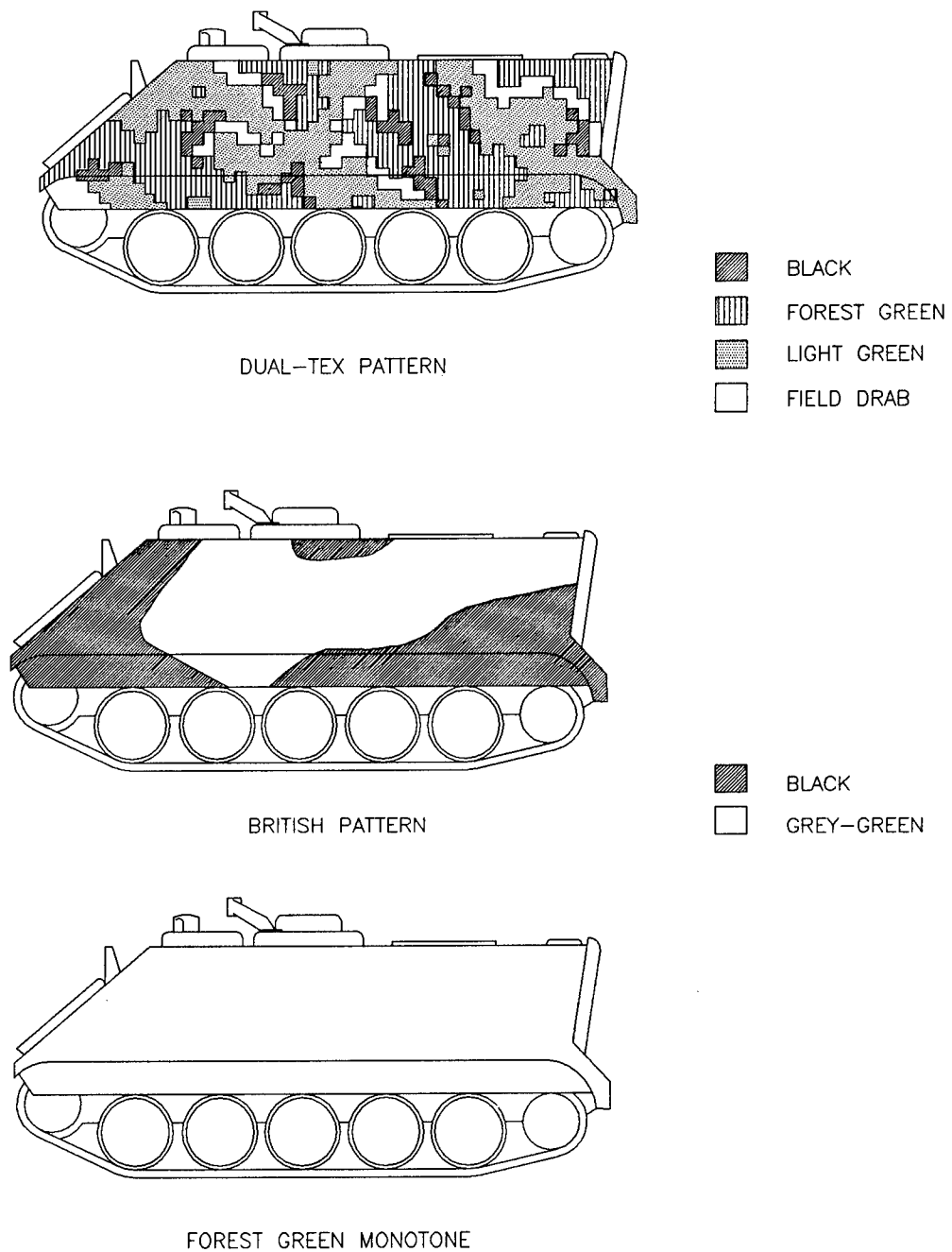
Figure 2 (Continued). Candidate Test Patterns/Color

straight edges and sharp angular features. About two-thirds of the vehicle was gray-green and the rest black. The paints used for this were glossy (18% reflectance), and were formulated to match British specifications.

E. Forest Green. The lusterless Forest Green M-113 represented the standard uncamouflaged monotone-painted vehicle.

3.2 Test Sites

The study was conducted at the Turner Drop Zone, Fort Devens, Massachusetts. This was a large cleared tract of land, surrounded by a mix of coniferous and deciduous forest resembling a central European background. Three test sites were selected. Site #1 was located on the western edge of the drop zone, so that the morning sun shone directly upon the target APC. Site #2 was located on the eastern edge of the drop zone, so that the afternoon sun shone directly upon the target. Site #3 was located at the northern end of the field. An observation path, starting at the opposite end of the drop zone from the target location, was laid out for each site. Each path followed zig-zag, random-length directions toward the test site, and afforded a continuous line of sight to its respective target location. The paths were surveyed and marked in approximately 50-meter intervals using random letter markers. Tables 1-3 show the markers and associated distances from the target for each of the three sites.

Table 1
Distance of Marker to Target APC for Site #1

Alphabet Marker	Distance in Meters from Marker to Target	Alphabet Marker	Distance in Meters from Marker to Target
C	1,173.70	F	608.16
U	1,132.02	Y	583.66
A	1,088.51	S	536.46
R	1,044.10	O	473.26
D	1,015.03	V	417.36
Z	989.27	G	373.08
H	947.17	L	315.16
B	901.17	N	269.62
K	854.06	I	215.97
P	808.71	Q	160.10
W	762.36	J	126.89
E	723.52	M	79.71
X	698.97	L'	27.65
T	653.54		

Table 2
Distance of Marker to Target APC for Site #2

Alphabet Marker	Distance in Meters from Marker to Target	Alphabet Marker	Distance in Meters from Marker to Target
Y	1,261.50	C	679.61
W	1,230.74	O	620.24
A	1,192.40	D	560.12
R	1,153.65	G	500.84
E	1,116.90	L	440.73
P	1,076.05	N	400.81
K	1,033.50	I	367.20
M	987.16	Q	328.03
J	942.80	U	279.77
T	902.04	H	225.50
Z	875.51	F	212.76
V	811.07	B	165.04
S	779.80	I'	108.60
X	737.09	J'	64.88

Table 3
Distance of Marker to Target APC for Site #3

Alphabet Marker	Distance in Meters from Marker to Target	Alphabet Marker	Distance in Meters from Marker to Target
T	1,189.64	X	560.12
P	1,141.91	L	548.76
V	1,080.25	N	513.50
F	1,027.07	I	473.84
H	973.92	E	413.48
S	931.76	Z	383.34
Q	884.10	G	323.84
A	827.81	O	281.60
U	778.91	J	225.49
K	719.76	M	175.06
C	679.61	W	120.84
Y	672.58	R	73.27
D	644.63	T'	23.84
B	593.39		

3.3 Test Subjects

A total of 204 enlisted soldiers from Fort Devens served as ground observers. All personnel had at least 20/30 corrected vision and normal color vision. A minimum of 32 observers were used for each test pattern/color, about evenly split between test sites. This number of

observers allowed the use of parametric statistics[1] in order to obtain definite conclusions from the analysis of the field data. Each observer was used only once.

3.4 Test Procedure

The test procedure was to determine the detection distances for each of the six test APCs, by searching for them while traveling along a predetermined measured path. Each ground observer started at the beginning of the observation path, i.e., marker C for Site #1, marker Y for Site #2, and marker T for Site #3. The observer rode in the back of an open 5/4-ton truck accompanied by a data collector. The truck traveled down the observation path at a very low speed, about 3-5 mph. The observer was instructed to look for military targets in all directions, except directly to his rear. When a possible target was detected, the observer informed the data collector and pointed to the target. The truck was immediately stopped, and the data collector sighted the apparent target. If the sighting was correct, i.e., an APC, the data collector recorded the alphabetical marker nearest the observation truck. If the detection was not correct, the data collector informed the observer to continue looking. The truck proceeded down the observation path. This search process was repeated, until the correct target was located. The painted APCs were rotated among the three sites, until each APC had at least ten observers at each site. Each APC was identically placed, so that a full side was facing the direction of observer approach.

4.0 Results

Tables 4 through 7 show the detection data for the five camouflage-patterned APCs and the monotone APC. Table 4 gives the mean detection range in meters for each test APC and its associated 95-percent confidence interval. Table 5 shows the analysis of variance performed upon the data of Table 4 to determine if there were significant differences in detection range, i.e., if the pattern/color of the APC had an effect upon detection range. Table 6 is the parametric test, Duncan's Multiple-Range, which indicates which test pattern/color differed significantly ($\alpha = 0.05$) from each other. Table 7 is the nonparametric statistical test[2] which complements the data of Table 6. Figure 3 is a graphic display of the detection ranges of Table 4.

ANITOLE, NEUBERT & JOHNSON

Table 4
Mean Detection Ranges (Meters) and 95-Percent
Confidence Intervals for Patterned/Color APCs

APC Pattern/Color	N	Mean	Standard Error	95-Percent Confidence Interval Lower Limit	Upper Limit
British	34	458.73	33.79	389.97	527.48
Forest Green	35	427.13	36.38	353.21	501.06
Swedish	37	455.18	37.47	379.20	531.16
Natural	34	249.35	21.00	206.62	292.09
U.S. Four-Color	32	480.29	50.87	376.53	584.05
Dual-Tex	32	364.68	23.22	317.33	412.03
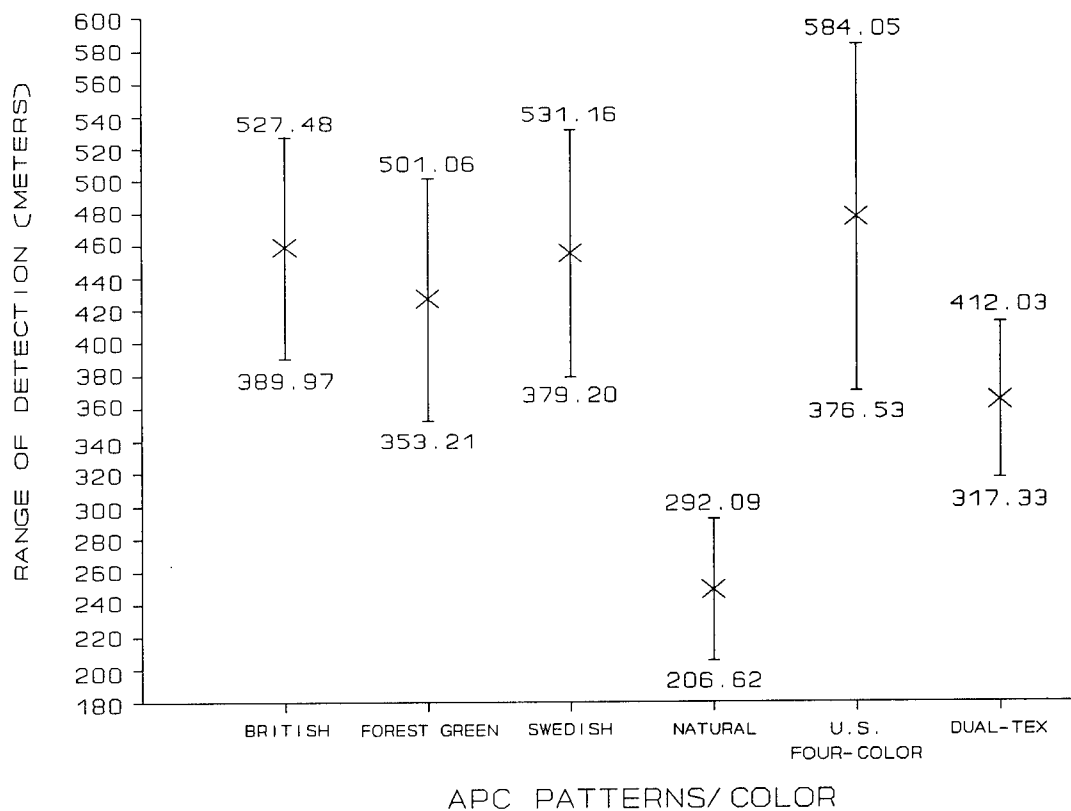
Figure 3. Mean Ranges of Detection
and 95-Percent Confidence Intervals for APC Patterns/Color

54

Table 5
Analysis of Variance for Patterned/Color APCs

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio	F Probability
Pattern/Color	5	1265130.04	253026.01	6.0194	0.000*
Error	198	8322951.82	42035.11		
Total	203	9588081.85			

Bartlett's Test for Homogeneous Variances
Number of Degrees of Freedom = 5
F = 6.388   Significance Level α = 0.000*
*Significant at α less than 0.001 level

Table 5 indicates that there were significant differences in the ability of the ground observers to detect the camouflage patterned/color APCs. The Bartlett's Test indicates that the variances for each level of camouflage pattern/color were not homogeneous, i.e., significantly different, so they are not necessarily from the same population.

Table 6
Duncan's Multiple-Range Test
Patterned/Color APCs (Range of Detection)

Best Subset 1		Subset 2		Worst Subset 3	
Pattern/Color	Mean	Pattern/Color	Mean	Pattern/Color	Mean
Natural	249.35	Dual-Tex	364.68	U.S. Four-Color	480.29
		Forest Green	427.13	Forest Green	427.13
		Swedish	455.18	Swedish	455.18
		British	458.73	British	458.73

The harmonic mean group size is 33.91. The subsets are significantly different at α = 0.05.

The Duncan's Multiple-Range Test is a parametric analysis that separates a set of significantly different means into subsets of homogeneous means. One of the assumptions is that each random sample is of equal size. Since that was not true, the harmonic mean of the group was used as the group size. As seen from the table, the Natural pattern was significantly best (α = 0.05) in having a mean range of detection of 249.35 meters, while the U.S. Four-Color pattern was among the worst, with a mean range of detection of 480.29 meters. The patterns/color Dual-Tex, Forest Green, Swedish, U.S. Four-Color, and British did not significantly differ from each other in mean ranges of detection.

Table 7 contains the nonparametric results of the Mann-Whitney Procedure, as a double check upon the Duncan's Test, in that the Bartlett's Test indicated that the variances for each level of camouflage pattern/color were not homogeneous. Parametric tests assume homogeneity of variance.

<div align="center">

Table 7
Mann-Whitney Procedure
Patterned/Color APCs (Range of Detection)

</div>

Best Subset 1 Pattern/Color	Mean	Worst Subset 2 Pattern/Color	Mean
Natural	249.35	Dual-Tex	364.68
		Forest Green	427.13
		Swedish	455.18
		British	458.73
		U.S. Four-Color	480.29

These results were very similar to those of Table 6. The Natural pattern was significantly ($\alpha = 0.05$) better than the remaining patterns. Note that the U.S. Four-Color pattern was not significantly different from the remaining patterns/colors.

5.0 Discussion

The Duncan's Multiple-Range Test (Table 6) and the Mann-Whitney Procedure (Table 7) were in close agreement with each other. Previous work by Neubert et al.[3] had found similar results. Even when the Bartlett's Test yielded a significant F value, the parametric Duncan's Multiple-Range Test yielded results consistent with those obtained from nonparametric tests. The Natural pattern, with a mean detection range of 249.35 meters, was significantly ($\alpha = 0.05$) more difficult to detect than the other patterns/color studied. This detection range was more than 100 meters shorter than its nearest competitor (Dual-Tex). The remaining patterns/color were very closely grouped with their mean ranges of detection between 427 and 480 meters.

6.0 Summary and Conclusions

A photographic process of developing a camouflage pattern analogous to the contrast highlights found in woodland backgrounds was presented, including the vehicle field test for evaluating the pattern. Five different camouflage-patterned APCs and one monotone APC were evaluated by

ground observers to determine their camouflage effectiveness as measured by range of detection. The following patterns/color were tested:

British Pattern
Forest Green Monotone
Swedish Pattern
Natural Pattern
U.S. Four-Color Pattern
Dual-Tex Pattern

A minimum of 32 military ground observers per pattern/color were used to enable the use of parametric statistics. This data was subjected to a statistical analysis using the methods of analysis of variance, the Duncan's Multiple-Range Test, and the Mann-Whitney Procedure. The analysis provided the following conclusions:

A. The Natural pattern had the shortest mean range of detection.
B. The Natural pattern was significantly ($\alpha = 0.05$) more difficult to detect than the other patterns/color.
C. The U.S. Four-Color pattern was the least effective in denying detection (Duncan's Multiple-Range Test).

7.0 Significant Accomplishment

Up until now, camouflage paint patterns have been subjectively developed based on the designer's artistic determination of shapes, pattern sizes, and colors. The photographic process described in this paper, that of determining the background highlights for use as camouflage patterns, is unique and captures the essence of the background better than any other method. This procedure presented an opportunity to scientifically derive a pattern.

The photographic pattern-generation procedure is also applicable to patterns for camouflage nets and uniforms for use in desert, arctic, and tropical backgrounds, as well as woodland. It is currently being considered for use to upgrade and optimize the patterns on the current production camouflage net, and the new ultralightweight camouflage net being investigated by Belvoir. U.S. patent #4,576,904 was issued to the principal investigator for this process.

8.0 References

1. Natrella, Mary G., *Experimental Statistics*, National Bureau of Standards Handbook 91, U.S. Department of Commerce, Washington, D.C., 1966.

2. Siegel, Sidney, *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Company Inc., 1956.

3. Neubert, Christopher J., Anitole, George, and Johnson, Ronald L., *Comparison of Parametric Versus Nonparametric Evaluation upon Non-Homogeneous Field Data*, Twenty-Eighth U.S. Army Operations Research Symposium, Fort Lee, Va., 10 October 1989.

Exertional Heatstroke in Soldiers: An Analysis of Recovery Rates, Predisposing Factors and Residual Heat Intolerance (U)

* Lawrence E. Armstrong, Dr., Jane·P. De Luca, Ms., Roger W. Hubbard, Dr.

U.S. Army Research Institute of Environmental Medicine, Heat Research Division, Natick, MA 01760-5007

INTRODUCTION

The risk of returning recovering heatstroke casualties to normal duty in a hot environment cannot be objectively determined at present. This research was designed to determine, by case histories, whether (1) certain individuals could remain heat intolerant for extended periods beyond clinical recovery from exertional heatstroke; (2) these same individuals could recover a "normal" state of heat acclimation, thereby ruling out a genetic predisposition to heat injury; and (3) diagnostic tests of physiological status could provide guidelines for therapeutic recovery.

Evaluating the ability of recovering prior heatstroke patients (PH) to exercise in a hot environment is a difficult process for attending physicians because these patients may be heat intolerant even though they have been declared "normal" in terms of clinical measurements. This can occur because the extent of the multiple tissue/organ injury of heatstroke and the rate of recovery from heatstroke are highly individualized. This individualized rate of recovery is largely due to the variabilty inherent in each episode (i.e. onset, duration, and intensity of hyperthermia) and to intra-subject variability. Keren et al.[1], for example, observed that the heat intolerance of one male, who experienced heatstroke twice, was temporary and was resolved five months after the second heatstroke episode. Bianchi et al.[2] found a variety of histological abnormalities in the liver tissue of two distance runners, which were resolved 11 - 12 months after heatstroke occurred. However, other laboratory trials[3,4] have demonstrated that some PH were heat intolerant 2 - 5 years after they experienced heatstroke.

Heat intolerance has been defined[3,5,6] as an inability to adapt to exercise in a hot environment, suggesting that classical heat acclimation adaptations (e.g. reductions in heart rate, rectal

temperature) do not occur in PH who are heat intolerant. Heat acclimation in healthy humans ordinarily requires 7 - 10 days and results in improved ability to exercise and live in the heat[7]. However, the nature and extent of the physiological adaptations of heat intolerant PH, heat tolerant PH, and normal individuals have not been clearly described. Although many case reports exist[1,8,9,10,11,12,13], only three previous controlled laboratory investigations[3,4,14] have utilized PH. However, none of these compared fluid-electrolyte, cardiovascular, or thermoregulatory responses of PH and control subjects during heat acclimation trials.

This investigation focused on the evaluation of 10 prior exertional heatstroke patients in our laboratory. The time elapsed between the exertional heatstroke episode and laboratory testing (mean $\pm$ SE: 61 $\pm$ 7 d) was a unique feature of this investigation, since other laboratory studies[3,4,14] evaluated PH 2 - 5 years after they had experienced heatstroke. The purposes of this investigation were to describe in PH the time course and nature of recovery from exertional heatstroke, their ability to acclimate to heat (daily 90 min exposure), and to identify the host or situational factors which may have predisposed PH to heatstroke.

We recognized that this information could be used to assess our current strategies for enhancing recovery, develop new physical training or heat acclimation programs if necessary, and provide objective data for accurate analysis of reinjury rates and risk. Also, data on the time course of recovery, assessed by clinical diagnostic tests of recovery and residual heat intolerance, could assist U.S Army commanders, training instructors and medical officers to determine the world-wide deployability of previously heat-injured soldiers.

METHODS

This investigation was approved by the USARIEM institutional review board, and informed, written consent was obtained from all subjects, in accordance with AR 70-25. Ten male PH participated in this study. PH were declared clinically normal by their attending physicians, completed a treadmill exercise stress test with no electrocardiographic abnormalities, and arrived at this laboratory 61 $\pm$ 7 d (range: 25 - 99 d) after exertional heatstroke; all PH were military personnel (4 officers, 6 enlisted). Medical records were obtained from the attending physicians. The criteria used to verify exertional heatstroke were : rectal temperature $\geq$ 40°C, altered mental status, and elevated serum creatinine phosphokinase (CPK), lactic dehydrogenase (LDH), aspartate aminotransferase (AST), and alanine aminotransferase (ALT).

Two PH (subjects A and G) had previously experienced heat exhaustion, but none had a history of heatstroke, childhood febrile seizures, or malignant hyperthermia. PH lived in both temperate and hot

ARMSTRONG, DE LUCA, HUBBARD

climates, prior to experiencing heatstroke. All mean physical characteristics of PH, except body fat %, were slightly above the mean of 1170 male soldiers described by De Luca et al.[15], and all fell within $\pm$ 1 standard deviation of the mean. The mean ($\pm$ SE) characteristics of PH were as follows: age - 26 $\pm$ 2 y (range: 21 - 44 y), height - 180 $\pm$ 3 cm, body mass - 85.420 $\pm$ 2.790 kg, surface area - 2.05 $\pm$ 0.04 m^2, mass-to-surface area ratio (M/SA) - 41.7 $\pm$ 0.7 $kg \cdot m^{-2}$, body fat % - 17.6 $\pm$ 1.3 %, maximal aerobic power (VO_2max) - 49.93 $\pm$ 2.05 $ml \cdot kg^{-1} \cdot min^{-1}$. A group of five healthy males served as a control (C) population, and undertook the same protocol. The mean ($\pm$ SE) characteristics of C were as follows: age - 25 $\pm$ 4 y (range: 18 - 42 y), height - 177 $\pm$ 4 cm, body mass - 78.250 $\pm$ 7.800 kg, surface area - 2.00 $\pm$ 0.09 m^2, M/SA - 39.7 $\pm$ 1.8 $kg \cdot m^{-2}$, body fat % - 14.7 $\pm$ 1.9 %, VO_2max - 52.11 $\pm$ 2.37 $ml \cdot kg^{-1} \cdot min^{-1}$. Physical characteristics of PH and C were not statistically different.

At the time of heatstroke, five PH were judged to be heat acclimatized, based on subject descriptions of activity and heat exposure. All subjects were unacclimatized at the onset of laboratory testing.

Seven days of heat acclimation were used to determine the heat tolerance of PH. All heat acclimation trials were conducted in an environmental chamber maintained at 40.1 $\pm$ 2.3^{0}C db, 23.8 $\pm$ 1.4^{0}C wb, and consisted of 90 min of treadmill walking (5.6 $km \cdot h^{-1}$, 5 % grade, 45 $\pm$ 2 % of VO_2max) for seven consecutive days. Although day 8 was the end point of heat acclimation, the data for day 7 are reported in Figure 1 because of circumstances beyond the investigators' control on day 8 (i.e. equipment malfunction). Maximal aerobic power (VO_2max) was measured on day 10 using a continuous treadmill test conducted at 20^{0}C.

Subjects were instructed to drink large quantities of water ad libitum before, during, and after each trial to insure adequate hydration, and were requested to produce a minimum daily urine volume of 2 1. A trial was terminated if heart rate (HR) exceeded 180 $beats \cdot min^{-1}$, if rectal temperature (T_{re}) exceeded 39.0^{0}C, or if T_{re} increased $\geq$ 0.6^{0}C during any five min period. HR was recorded continuously using an electrocardiographic telemetry system (Hewlett-Packard). T_{re} was measured via a rectal probe inserted 8 cm beyond the anal sphincter, and mean weighted skin temperature (T_{sk}) was calculated by using a three site derivation. The rate of body heat storage was calculated using the following formula: Heat Storage ($Cal \cdot m^{-2} \cdot h^{-1}$) = $M \cdot S \cdot MBT \cdot SA^{-1}$ (Eq. 1); where M = body mass (kg), S = specific heat constant (0.83 $W \cdot h^{-1} \cdot kg^{-1} \cdot {}^0C^{-1}$), MBT = change in mean body temperature calculated as (0.8 $\cdot T_{re}$ + 0.2 $\cdot T_{sk}$), and SA = surface area (m^2)[16].

Whole body sweat rate ($1 \cdot h^{-1}$) was measured using body mass differences (corrected for water intake and urine output) pre-post trial, and was normalized for surface area ($g \cdot m^{-2} \cdot h^{-1}$). Sweat sensitivity was calculated as a measure of sweat rate per degree rise in T_{re}

ARMSTRONG, DE LUCA, HUBBARD

$(g \cdot m^{-2} \cdot h^{-1} \cdot {}^0C^{-1})$. The number of heat activated sweat glands was determined at the end of exercise on day 1 of heat acclimation, by covering an area of skin over the scapula with a layer of vaseline petroleum jelly. Microphotography (35mm) was used to provide a permanent record of the number of active sweat glands (appearing as a bead of sweat) per cm^2. Sweat electrolyte losses during exercise were measured on days 2, 5, 8 using a previously described whole body washdown technique[17,18]. Sweat electrolyte loss (total mEq) was calculated by multiplying the volume of rinse water by the concentration of electrolytes in the rinse water. Sweat electrolyte concentration $(mEq \cdot l^{-1})$ was calculated using total electrolyte loss and whole body sweat rate.
Pre-exercise body mass and urine specific gravity were measured each day at 0730h.

Urine volume and electrolyte concentration $(mEq \cdot l^{-1})$ were used to calculate the daily loss of electrolytes in the urine $(mEq \cdot 24\ h^{-1})$. Subjects carried clean, inert urine collection containers during their daily business and returned the containers every 24 h. Sodium (Na^+), potassium (K^+), and caloric intake $(mEq \cdot d^{-1})$ were determined from dietary records which subjects maintained throughout the investigation.

Blood samples were taken from an antecubital vein pre-exercise, after subjects stood quietly for 20 min in the heat, and post-exercise. The following clinical assays were performed: hematocrit (standard microcapillary technique), hemoglobin (cyanmethemoglobin technique), total protein, osmolality (freezing point depression), and the electrolytes Na^+ and K^+ (flame photometry). Percentage change in plasma volume (PV%) was calculated from pre-exercise hematocrit and hemoglobin values[19]. Blood samples were were also analyzed for the following clinically relevant enzymes: CPK, LDH, AST, and ALT. If a subject had enzyme levels above the normal clinical range, he was scheduled for a second iteration of this protocol within 3 - 6 months.

Two-way ANOVA with Newman-Keuls post hoc comparisons were used to identify significant differences between PH and C, and between heat acclimation days. Student's t-test was used to compare descriptive characteristics, training, and VO_2submax between groups. Statistical correlation coefficients were calculated via multiple linear regression analysis. It should be noted that all subjects underwent a 15 min step test[4] in a temperate environment (25.8^0C), two days prior to the start of heat acclimation; those results have been published elsewhere[20].

RESULTS AND DISCUSSION

Table 1 lists selected characteristics of PH during the heatstroke episode and subsequent hospitalization. The duration of altered mental status was not significantly correlated ($p > .05$) with either the peak serum enzyme levels during hospitalization (Table 1) or the peak serum

Table 1. Characteristics of PH during field observations and hospitalization.

Subject	Maximal T_{re} (°C)*	Spontaneous cooling?	Mental status and duration (h)	Peak post-heatstroke serum enzyme levels **			
				CPK (units·l^{-1})	LDH (units·l^{-1})	AST (units·l^{-1})	ALT (units·l^{-1})
A	40.0	no	disoriented (---)	18,120 (5)	500 (4)	281 (5)	196 (3)
B	41.1	no	coma (7.5)	---	***	***	---
C	41.1	no	coma (0.4)	263 (1)	294 (1)	351 (3)	221 (5)
D	41.1	no	disoriented (0.3)	327 (2)	821 (2)	4,680 (2)	---
E	41.1	no	disoriented (1.0)	1,703 (1)	405 (1)	149 (2)	148 (2)
F	41.1	---	disoriented (5.0)	1,928 (1)	406 (1)	111 (4)	202 (5)
G	41.1	---	coma (0.4)	3,820 (1)	---	206 (2)	210 (3)
H	41.8	no	coma (---)	7,480 (1)	522 (1)	284 (2)	407 (2)
I	41.1	---	coma (0.1)	3,182 (1)	260 (1)	---	88 (1)
J	40.4	no	disoriented (0.2)	14,160 (2)	1,390 (2)	1,200 (2)	2,145 (3)
mean	41.0			5,664	575	908	452
± SE	±0.2			± 2,133	± 132	± 553	± 294

* - All measurements were taken in the field, except patient J (emergency room). Some rectal thermometers read only to 41.1°C.

** - Number of days to reach maximal level is shown in parentheses. Normal enzyme concentration ranges are: CPK – 33–213 units·l^{-1}, LDH – 92–186 units·l^{-1}, AST – 7–32 units·l^{-1}, ALT – 2–45 units·l^{-1}.

*** - Serum enzyme measurements were made on post-heatstroke day 4 only, as follows: LDH – 865 units·l^{-1}, AST – 490 units·l^{-1}.

ARMSTRONG, DE LUCA, HUBBARD

enzyme levels observed during heat acclimation trials. Sleep loss, generalized fatigue, a long exercise bout (or many exercise sessions each day), and a single long heat exposure were the most prevalent predisposing factors acknowledged by PH during a pre-study interview.

The comparison of the heat acclimation data (e.g. HR and T_{re}) of PH and C (Fig. 1) demonstrated that nine PH and all C were physiologically normal. Both PH and C showed significant decreases in HR ($p < .05$) and T_{re} ($p < .025$) by day 7. Therefore, the designation of "clinically normal" for PH at the treatment facility was equivalent to "heat tolerant" in our climatic chamber, at 61 ± 7 d after exertional heatstroke (range: 25 - 99 d). Other thermoregulatory measurements made during 90 min heat acclimation trials, for the nine heat tolerant PH and C, indicated no significant between-day differences in the following variables: T_{sk}, heat storage, $T_{re} - T_{sk}$, heat activated sweat glands, sweat rate, and sweat sensitivity. Previous investigations[21] indicate that significant between-day differences (e.g. sweat rate, sweat sensitivity, T_{sk}) would have been observed had PH and C been exposed to heat for a longer period (i.e 14 d). It was concluded that no substantial sodium or potassium deficit (involving electrolyte losses in sweat, urine, and feces[22]) occurred in either PH or C, as a result of daily 90 min trials. There were no between-day or between-group differences in mean pre-exercise body mass, 24 h caloric intake, 24 h urine volume, or urine specific gravity. All mean urine specific gravity values were < 1.020 on all days, for PH and C. There were no significant differences between PH and C in any of the following blood measurements: hematocrit, hemoglobin, total protein, Na+, K+, osmolality. The between-day PV% were positive and large in both groups by day 5 (PH: $+14.9 \pm 2.9\%$; C: $+17.1 \pm 3.1\%$).

Subject A was defined as heat intolerant, in spite of normal sodium/potassium balance and clinical indices, using the definition of Strydom[23] and others[3,5], since he was unable ($T_{re} > 39^{0}C$) to adapt to exercise during 90 min trials. Subject A was scheduled for a second and third iteration of this protocol, at seven and 11.5 months following heatstroke, due to his inability to acclimate. He was heat intolerant during the second iteration, but exhibited improved HR and T_{re} adaptations during the third iteration, and was defined heat tolerant at that time. The characteristics for which subject A was at the extreme of all PH and C were: body fat %, age, VO_2max, two previous experiences with heat exhaustion, the peak post-heatstroke CPK level (Table 1), and the number of days to reach peak post-heatstroke CPK, ALT and AST levels (Table 1). He exhibited no signs of low-grade viral or bacterial infections. His responses were similar to the heat intolerant PH described by Shapiro[14] and Shvartz[4]. Earlier it had been reported[2] that the recovery from exertional heatstroke-induced hepatic injury in two distance runners was not complete until 11 - 12 months after heatstroke. Although this lengthy recovery time agrees well with the recovery of

ARMSTRONG, DE LUCA, HUBBARD

Figure 1. Mean ($\pm$ SE) final HR and T_{re} values during heat

acclimation, on days 2 and 7. Symbols: solid bars —

PH, shaded bars — C, solid circles — subject A.
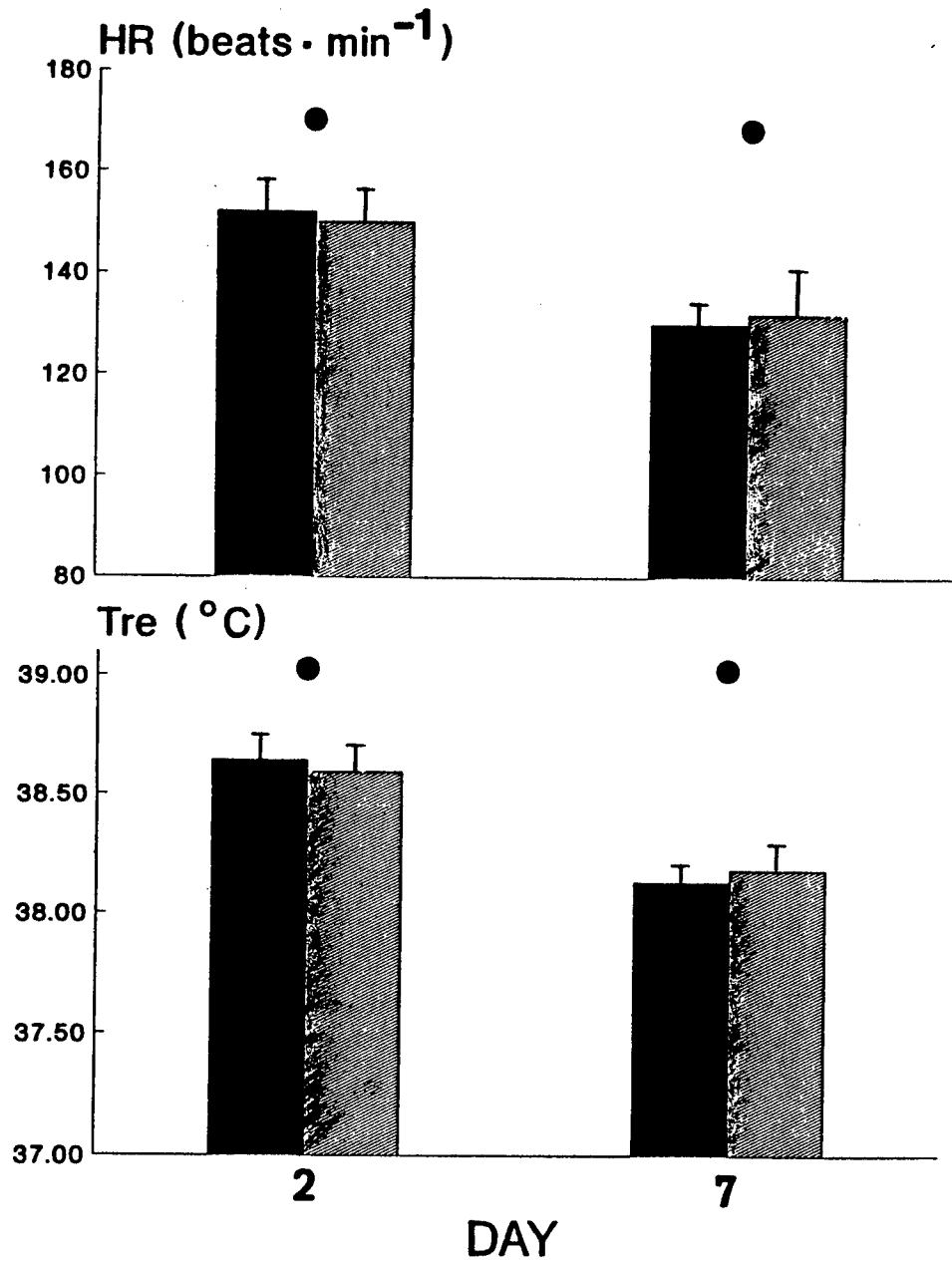
ARMSTRONG, DE LUCA, HUBBARD

heat acclimation ability in subject A (at 11.5 months), this does not indicate that he had severe liver damage.

Senay and Kok[5] described heat intolerant miners by their inability to acclimate to repeated days of exercise in hot, humid environments. Wyndham[13] and Strydom[6] reported that two to five per cent of these recruits were innately heat intolerant. At the onset of the current investigation, it was not known whether PH were members of such a heat intolerant group, or whether situational/host factors predisposed them to heatstroke. Because nine PH exhibited HR and T_{re} responses that were statistically similar to those of C (fig. 1), hereditary heat intolerance was excluded as a causative factor in all subjects except subject A during iteration 1. The fact that subject A ultimately acclimated to exercise in the heat, at 11.5 months after heatstroke (iteration 3), suggests: (a) he was not hereditarily, permanently heat intolerant prior to heatstroke; (b) the concept that "one heatstroke predisposes to another heatstroke" may be true temporarily, and was more likely to be true during the time that subject A was heat intolerant[1]; (c) some physiological factor(s) changed to allow subject A to respond normally during iteration 3.

Serum CPK levels indicated that three PH (subjects E, F, G) had high CPK values (838, 5625, 958 $U \cdot l^{-1}$, respectively; normal range: 33 – 213 $U \cdot l^{-1}$), and they were asked to return for additional testing. They were treated separately (as group P2) from the other seven PH (group P1) during enzyme data analysis (Fig. 2). Isoenzyme assays indicated that the CPK in the serum of P2 originated in muscle tissue (> 98 % m-m band), in all cases. These CPK elevations during exercise-heat tolerance tests are probably explained by the inactivity and detraining of P2 following heatstroke. In fact, changes in serum CPK, LDH, AST, and ALT during heat acclimation did not follow the course of serum enzyme changes after heatstroke (Table 1), suggesting a response to exercise rather than a pathological condition. It has been demonstrated[23] that similar CPK elevations occurred following a single eccentric exercise trial.

Only one research team has previously reported the serum enzymes of PH at any time greater than two weeks following exertional heatstroke[9], but their data indicated little about the time course of recovery. Figure 3 depicts the serum enzyme levels of subject F and illustrates the trend which was noted in all members of group P2. This trend involved reductions of serum CPK, LDH, AST, and ALT, from iteration 1 to iteration 2 or 3. This trend hypothetically may be explained by the restoration of normal membrane/organ function, increased physical training between successive iterations, or combinations of these factors.

ARMSTRONG, DE LUCA, HUBBARD

Figure 2. Pre-exercise serum CPK, LDH, AST and ALT values (mean $\pm$ SE) of groups P1 (n = 7), P2 (n = 3), and Control (n = 5).
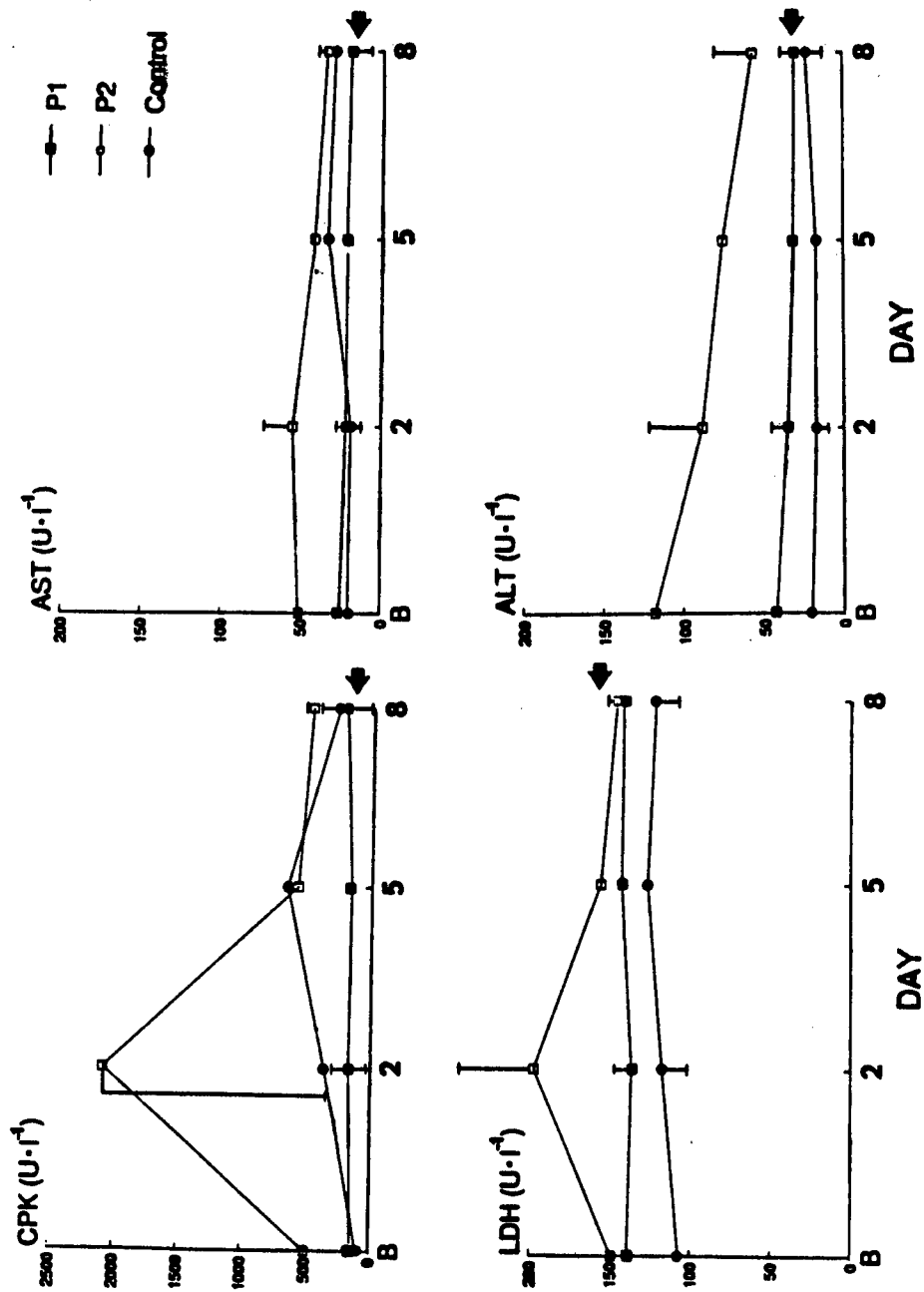
Figure 3.   Pre-exercise serum CPK, LDH, AST and ALT values of subject F.
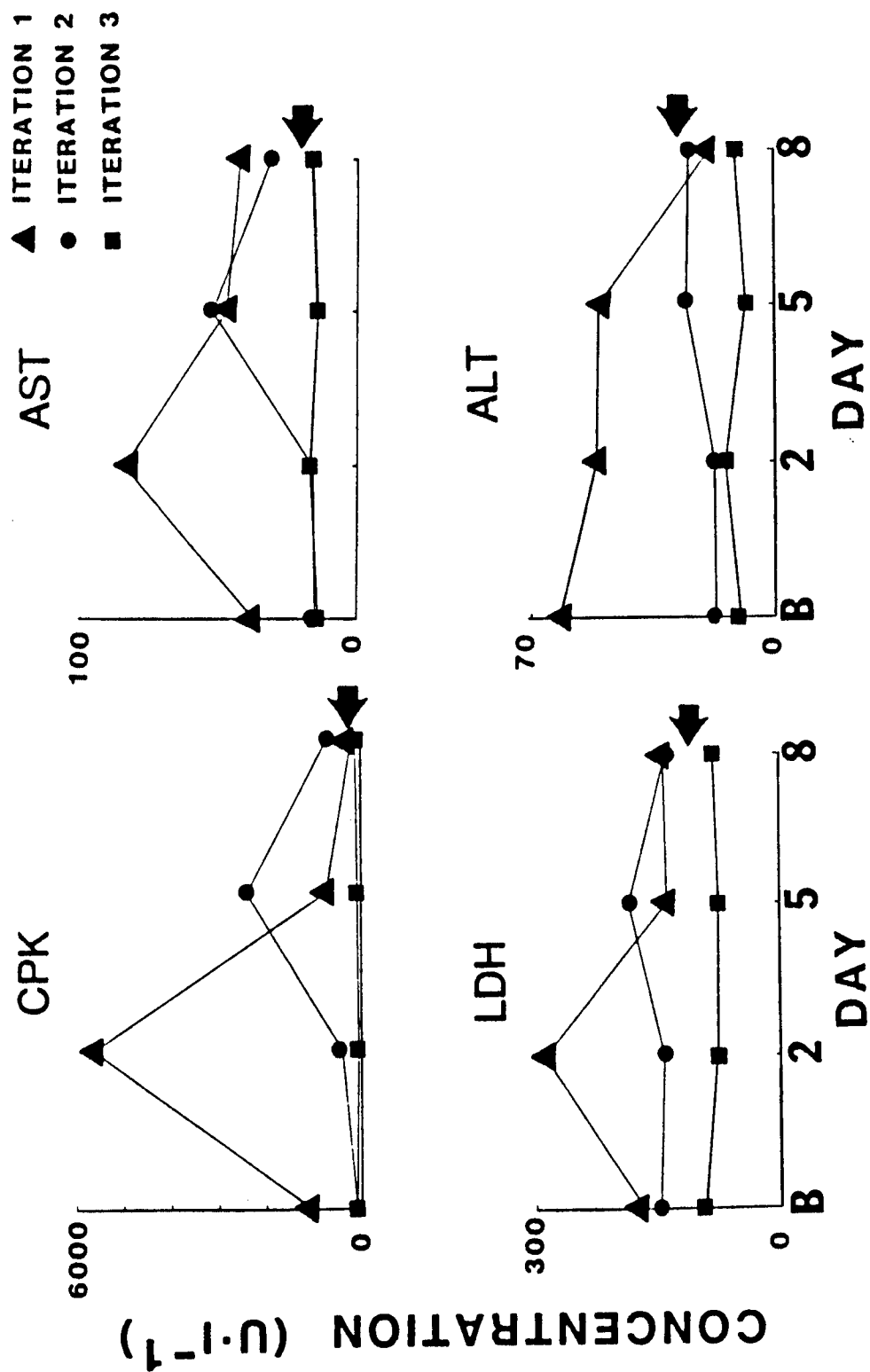
ARMSTRONG, DE LUCA, HUBBARD

CONCLUSIONS

1. At 61 ± 7 d following heatstroke, nine out of 10 PH exhibited normal heat acclimation, thermoregulation, whole body sodium and potassium balance, sweat gland function, and clinical values. Thus, the current recommendation to restrict all physical activity during recovery can be counter-productive.

2. Only subject A was defined heat intolerant (at approximately two and seven months after heatstroke), but was defined heat tolerant at 11.5 months, demonstrating that complete recovery from exertional heatstroke may require 11.5 months.

3. None of the 10 PH in this investigation were hereditarily heat intolerant, based upon their ability to acclimate to heat. Multiple stressors (e.g. sleep loss, sudden increase in physical training, lengthy exposure to heat stress) may be critical in inducing heat injury.

4. Of the factors measured in the hospital or in the laboratory, none was clearly related to prognosis, recovery from heatstroke, or performance during heat acclimation trials. Previously described prognostic factors[24] (e.g. AST > 1000 U·1^{-1}, coma duration > 2 h) were not predictive of the ability of PH to work in the heat.

5. The results of the current investigation indicate that clinically normal PH, during recovery, can benefit from well-planned physical training and heat acclimation programs which include on-site medical monitoring, gradual increments in the duration and intensity of heat exposure (e.g. core body temperature increase expressed in terms of degree-minutes[25]), and gradual increments in exercise frequency/duration/intensity.

REFERENCES

1. Keren, G., Y. Epstein, and A. Magazanik. Temporary heat intolerance in a heatstroke patient. Aviat. Space Environ. Med. 52:116-117,1981.
2. Bianchi, L., H. Ohnacker, K. Beck, and M. Zimmerli-Ning. Liver damage in heat stroke and its regression. Hum. Pathol. 3:237-249, 1972.
3. Epstein, Y., Y. Shapiro, and S. Brill. Role of surface area-to-mass ratio and work efficiency in heat tolerance. J. Appl. Physiol. 54:831-836, 1983.
4. Shvartz, E., S. Shibolet, A. Merez, A. Magazanik, and Y. Shapiro. Prediction of heat tolerance from heart rate and rectal temperature in a temperate environment. J. Appl. Physiol. 43:684-688, 1977.
5. Senay, L.C. and R. Kok. Body fluid responses of heat-tolerant and intolerant men to work in a hot wet environment. J. Appl. Physiol. 40:55-59, 1976.
6. Strydom, N.B. Heat intolerance: its detection and elimination in the mining industry. S. Afr. J. Sci. 76:154-156, 1980.

7. Wenger, C.B. Human heat acclimatization. In: Human Performance Physiology and Environmental Medicine at Terrestrial Extremes, K.B. Pandolf, M.N. Sawka, and R.R. Gonzalez (Eds.). Indianapolis: Benchmark Press, 1988, pp. 153-197.

8. Assia, E., Y. Epstein, and Y. Shapiro. Fatal heat stroke after a short march at night: a case report. Aviat. Space Environ. Med. 56:441-442, 1985.

9. Beard, M.E.J., J.W. Hamer, G. Hamilton, and A.H. Maslowski. Jogger's heat stroke. N. Z. Med. J. 89:159-161, 1979.

10. Dukes-DoBose, F.N. Hazards of heat exposure: a review. Scand. J. WorkEnviron. Health 7:73-83, 1981.

11. Robinson, S., S.L. Wiley, L.G. Myhre, S. Bondurant, and J.J. Mamlin. Temperature regulation of men following heatstroke. Isr. J. Med. Sci. 12:786-795, 1976.

12. Sohar, E., D. Michaeli, U. Waks, and S. Shibolet. Heatstroke caused by dehydration and physical effort. Arch. Intern. Med. 122:159-161, 1968.

13. Wyndham, C.H. The physiology of exercise under heat stress. Ann. Rev. Physiol. 35:193-220, 1973.

14. Shapiro, Y., A. Magazanik, R. Udassin, G. Ben-Baruch, E. Shvartz, and Y. Shoenfeld. Heat intolerance in former heatstroke patients. Ann. Intern. Med. 90:913-916, 1979.

15. De Luca, J.P., L.E. Armstrong, E.L. Christensen, R.W. Hubbard, J.A. Vogel, and D.D. Schnakenberg. Mass-to-surface area ratio in military personnel. Natick, MA: U.S. Army Research Institute of Environmental Medicine,Technical Report No. T21-88, 1988, pp. 1-37.

16. Craig, F.N., H.W. Garren, H. Frankel, and V.W. Blevins. Heat load and voluntary tolerance time. J. Appl. Physiol. 6:634-644, 1954.

17. Armstrong, L.E., R.W. Hubbard, J.P. De Luca, and E.L. Christensen. Heat acclimatization during summer running in the northeastern United States. Med. Sci. Sports Exerc. 19:131-136, 1987.

18. Armstrong, L.E., R.W. Hubbard, P.C. Szlyk, W.T. Matthew, and I.V. Sils. Voluntary dehydration and electrolyte losses during prolonged exercise in the heat. Aviat. Space Environ. Med. 56:765-770, 1985.

19. Dill, D.B. and D.L. Costill. Calculation of percentage changes in volumes of blood, plasma, and red cells in dehydration. J. Appl. Physiol. 37:247-248, 1974.

20. Armstrong, L.E., R.W. Hubbard, J.P. DeLuca, E.L Christensen, and W.J. Kraemer. Evaluation of a temperate environment test of heat tolerance in prior heatstroke patients and controls. Eur. J. Appl.Physiol., in press, 1990.

21. Wyndham, C.H., A.J.A. Benade, C.G. Williams, N.B. Strydom, A. Golden, and A.J.A. Heynes. Changes in central circulation and body fluid spaces during acclimatization to heat. J. Appl. Physiol. 25:586-593, 1968.

ARMSTRONG, DE LUCA, HUBBARD

22. Lentner, C. (Ed.) <u>Geigy Scientific Tables.</u> West Caldwell, NJ: Ciba-Geigy Corporation: 1981, pp. 151-158.
23. Newham, D.J., D.A. Jones, and R.H.T. Edwards. Large delayed plasma creatinine kinase changes after stepping exercise. <u>Muscle Nerve</u> 6:380-385, 1983.
24. Shibolet, S., M.C. Lancaster, and Y. Danon, Heatstroke: a review. <u>Aviat. Space Environ. Med.</u> 47:280-301, 1976.
25. Hubbard, R.W., C.B. Matthew, M.J. Durkot, and R.P. Francesconi. Novel approaches to the pathophysiology of heatstroke: the energy depletion model. <u>Ann. Emerg. Med.</u> 16:1066-1075, 1987.

An Inactivated Whole Virus Vaccine for the Prevention of Viral Hepatitis, Type A (U)

*William H. Bancroft, COL, Leonard N. Binn, Dr.,
Kenneth H. Eckels, Dr., James W. LeDuc, LTC,
Maria H. Sjogren, LTC, Doria R. Dubois, Dr.,
Ruth H. Marchwicki, M.S., and Charles H. Hoke, COL
Walter Reed Army Institute of Research
Washington, D.C. 20307-5100;
United States Army Medical Research Institute
of Infectious Disease
Frederick, Maryland 21701-5012

HEPATITIS A AS A MILITARY THREAT

Although acute hepatitis has been recognized as a military health problem since the American Civil War, an understanding of the etiology, means of transmission and methods for prevention only began to emerge during WWII (1). Researchers sponsored by the Surgeon General of the Army made systematic observations of natural and experimental infections. They determined that one type of hepatitis was shed in human feces and could be transmitted by contaminated water. Following recovery from "infectious hepatitis", a patient was solidly immune to re-infection. These observations led directly to the chlorination of water to inactivate the infectious agent and recommendations for the prophylactic administration of large intramuscular doses of gamma globulin to prevent disease (2). These achievements were major contributions to improved public health, but there has been no fundamental advance in the prevention of hepatitis A in the subsequent 45 years.

In Vietnam, the US Army found acute hepatitis to be a major medical cause of time lost from duty (3). Patients with viral hepatitis averaged 35 days in the hospital. Hepatitis caused over 80,000 man days lost from duty each year, peaking at 116,000 man days in 1968. Routine serology for the diagnosis of hepatitis was not available then, but it is believed that nearly all cases were hepatitis types A and B.

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

During the past 9 years, four studies (4-7) have been published on the etiology of viral hepatitis seen in US Army hospitals (Table 1). In each report, Hepatitis B is the most common cause of disease during peacetime. Hepatitis A is more common in soldiers assigned to Korea than to the US or Europe. Taken together, Hepatitis A and B account for 2/3 or more of all cases of viral hepatitis seen in US Army hospitals. Prevention of both Hepatitis A and B by vaccination should significantly reduce the morbidity from hepatitis in American forces. Highly effective hepatitis B vaccines are already licensed.

TABLE 1. ETIOLOGY OF ACUTE HEPATITIS
U.S. ARMY
1972-1978

	Hepatitis A	Hepatitis B	NonA,NonB
Asia	15-25%	75-80%	0-3%
Europe	1-5	69-72	17-27
United States	13	50	37

Hepatitis A is notorious as a food-borne or water-borne threat. An example is the 1974 epidemic attributed to an infected foodhandler working in a mess hall at the USN Training Center in San Diego (8). With an attack rate of 47.8 per 1000 at risk, this epidemic involved 133 cases in only one month. This common source outbreak of hepatitis A struck many people suddenly, but was not sustained by continued viral transmission.

Hepatitis A outbreaks are increasingly frequent among children attending Child Care Centers and their parents. An outbreak at the Child Care Center at Fort Richardson, Alaska involved 119 people over a 9 month period (9). In this setting, symptomatic children can shed large amounts of HAV and secondary virus transmission is easily sustained by the frequent introduction of new susceptible children and adults into the infected environment. Sustained viral transmission to newly introduced susceptibles occurred in allied forces in North Africa during WWII (1).

Since 1978 there has been a steady decline in the number of cases of hepatitis A admitted to US Army hospitals in all regions (ET Takafuji, personal communication). The decline parallels a downward trend in reported cases throughout the United States in recent years (10). The falling hepatitis rates within this country must not lead to complacency of travelers.

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

TABLE 2. HEPATITIS A ANTIBODY IN
U.S. MILITARY PERSONNEL

Year	Group	Hepatitis A Positive	
		No.	(%)
1973	Army/USAF	176	22
1978	Special Forces	375	22
1981	Marines	318	26
1981	Army Airborne	233	15
1983	Armored Cavalry	413	22
1989	Army Airborne	274	13

Table 2 shows there also has been a slow decline in the prevalence of hepatitis A antibody in American soldiers (1). More recently, a random sample of 1740 new US Army recruits in 1989 found only 8.9% to be seropositive (Kadlec, P. Kelley, M. Sjogren, unpublished observation). The inescapable conclusion is that American service personnel are even more vulnerable to hepatitis A infection now than they were during the Vietnam era.

HUMAN HEPATITIS A INFECTIONS

Although children frequently have mild or totally asymptomatic infections with hepatitis A virus, Table 3 shows that 76-97% of infected American soldiers become acutely ill and unable to perform their usual activities (11,13,14). The onset of acute symptoms is preceeded by the appearance of HAV in the stool (12). Virus is detectable in the cytoplasm of hepatocytes (13). At the onset of symptoms, IgM antibody to HAV is detectable in the serum (14), leading some investigators to conclude that the pathogenesis of hepatitis involves the immune response to the virus (12). This is supported by the observation that human lymphocytes having the phenotype of Natural Killer Cells selectively lyse hepatitis A infected cells and release interferon alpha (15).

Patients with hepatitis A can be expected to recover completely within a few weeks. Mortality is estimated to be <0.1%. Although chronic carriers are not known to occur, Sjogren et al (16) found that 6.6% young adults recovering from acute hepatitis A in Argentina demonstrated a relapse of clinical hepatitis between 30 and 90 days after the initial episode accompanied by recurrent viral shedding in the feces.

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

TABLE 3. ACUTE HEPATITIS A IN AMERICAN SOLDIERS

Any Symptoms	76-97%
Dark Urine	67-97
Nausea/Vomiting	70-83
Yellow Eyes	41-76
Abdominal Pain	57-66
Malaise	52-79
Fever	48-72
Mortality	< 0.1%

CURRENT MEANS OF PREVENTING HEPATITIS A

The prevention of acute hepatitis A depends on good routine personal hygiene to prevent fecal-oral virus transmission. The administration of immune globulin (IG) in doses of 0.02-0.06 ml/kg is advised for pre-exposure protection of travelers to endemic areas (17). Sustained protection requires reinjection of IG every 5 months. For postexposure protection, a single dose of 0.02 ml/kg is advised. The average American soldier deployed to an endemic area receives 5 ml of IG. The large volume injections are painful and compliance with recommendations for readministration is often inadequate. In addition, the administration of IG during military operations in endemic areas poses a potentially avoidable problem for logistics. Recently, American lots of IG have shown declining levels of hepatitis A antibody, corresponding to the fall in disease rates nationwide (17).

Unfortunately, hepatitis A remains a threat. With the exception of the Scandinavian countries, all other geographic regions have more hepatitis A infections than North America. Susceptible travelers to regions of poor sanitation remain at risk. A vaccine to provide long-lasting protection against acute hepatitis A would replace the use of IG and benefit military personnel, other travelers, child care attendees and health care workers.

HEPATITIS A VIRUS

Hepatitis A virus (HAV) is a member of the Picornaviridae family by virtue of its biophysical characteristics, the organization of its single stranded ribonucleic acid and the functions of its proteins. Picornaviruses include the following four established genera and representative viruses: Enterovirus

(Poliovirus), Rhinovirus (common cold virus), Apthovirus (Foot and Mouth Disease Virus) and Cardiovirus (Encephalomyocarditis Virus). HAV differs from these genera in having only a single serotype, specific trophism for liver cells in the infected host, and little nucleotide and amino acid homology with other picornaviruses. Of importance to vaccine development, field strains of HAV replicate very slowly in cell culture, yield low levels of infectious virus and do not cause cytopathic effect, unlike other picornaviruses. It is believed that HAV belongs to a unique genus yet to be named (18).

EARLY EFFORTS AT VACCINE DEVELOPMENT

Provost et al (19) first demonstrated the feasibility of protective immunization by administering an extract of infected marmoset liver inactivated with formalin to other marmosets and challenging them with live virus. Relatively low virus yields from cell culture led manufacturers to prepare live, attenuated vaccines or inactivated subunit vaccines produced by recombinant technology rather than inactivated whole virus (20,21). Elevated serum transaminases in some of the first recipients and the possibility for reversion of virulence discouraged further development of live products (22). The recombinant products were no more successful for hepatitis A than similar attempts for poliovirus. Although these approaches may ultimately succeed, both types of vaccine face difficulties in meeting federal regulatory standards.

The authors began working of Hepatitis A in 1976 and, in 1983, decided to prepare an inactivated whole virus vaccine derived from cell culture as an alternative to the commercial approaches. The overall objectives of the program are:

1. To select safe, effective and affordable hepatitis vaccines.
2. Prevent the two most common causes of acute hepatitis in the military, hepatitis A and B.
3. Routine immunization of travelers to endemic areas, health care personnel and, if needed, all military recruits.

The initial step was a study of the host range of HAV, in primates and primate cell cultures. Binn et al (23) found African Green Monkey Kidney (AGMK) cells to be best for primary isolation of HAV. A continuous line of AGMK cells (BSC-1) allowed passaged virus to grow to the highest titer. Infected BSC-1 cells were detected by immune autoradiography with the

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

invention of a novel Radioimmunofocus Assay (RIFA) for precisely quantifying the infectivity of a non-cytopathic virus (24). RIFA identifies infected cells in vitro by holding the infected cell monolayer under an agar overlay for 14 days, then flooding the uncovered cells with radioiodinated immune IgG, washing off uncombined antibody and allowing the cells to expose Xray film. Radioactive antibody complexed to Hepatitis A antigen creates dark spots on the Xray film which can be counted as in a standard plaque assay for precise quantitation of infectious virus particles. The test can also be modified to allow precise quantitation of neutralizing antibody (Radioimmunofocus Inhibition Test) as in a conventional plaque reduction neutralization test (25). Indirect evidence from human recipients of IG indicates that serum neutralizing antibody titers of 10-40 measured by the radioimmunofocus inhibition test are protective for humans (26).

Careful screening of primate sera for hepatitis A antibody led to the discovery that Panamanian Owl Monkeys, *Aotus trivirgatus*, are susceptible to natural infection with HAV (27). Owl monkeys can be infected orally or by injection and are useful experimental models of human hepatitis A infection for studies of pathogenesis and vaccine efficacy (28,29). This original observation is important because owl monkeys are more easily obtained by American investigators than are marmosets and chimpanzees, the more commonly used primates.

WRAIR HEPATITIS A VACCINE

A prototype inactivated hepatitis A vaccine was prepared by harvesting the HM-175 strain of HAV from infected BSC-1 cells and inactivating the virus with formalin for 12 days (30). Six monkeys received 3 doses of vaccine at one month intervals. All animals developed serum antibody by radioimmunoassay and the neutralization test following the first two doses. The peak gometric mean titer (GMT) of neutralizing antibody was 320 at 84 days after the first dose. Eleven weeks following the third dose, all immunized monkeys and 3 seronegative controls were challenged with 10^6 monkey infectious doses of wild hepatitis A virus given either orally or intravenously. All three control animals developed acute hepatitis with fecal shedding of viral antigen followed by the development of serum antibody. None of the immunized animals shed fecal antigen nor developed hepatitis despite the high challenge dose. The possibility of preparing a vaccine for human use seemed likely, but required modification of the production methodology.

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

TABLE 4. HEPATITIS A HM-175/FI-1 VACCINE
PREPARATION AND CHARACTERISTICS

Parent Virus
  HM-175 strain - Human isolate from Australia (1976)

Passage History
  Marmoset liver (NIAID) - 6 times
  African green monkey kidney cells (AGMK) - 10 times
  Human diploid lung cells (MRC-5) - 7 times

Infected Cell Pool
  MRC-5 passage 7 - Disrupted cells on Day 42
  Clarification - Centrifuged at 8000 rpm for 30 min
  Infectivity titer - 8.2 X 10^6 RFU/ml
  Pre-inactivation filtration - 0.2 micron

Inactivation
  Formalin - 0.05% at 35° for 12 days
  Post-inactivation filtration on Day 5 - 0.2 micron
  Infectivity titer - 0 RFU on Days 5 and 12
  Infectivity in vivo - 0/3 owl monkeys

Final Product
  Storage - 10 ml/vial at 4°
  HAV antigen - 2-17 ng/ml (3 independent estimates)
  Total protein - 100 mcg/ml
  Formalin - 0.035%
  Stability - Mouse potency maintained 5 years

The first inactivated hepatitis A vaccine administered to humans was prepared with the HM-175 strain of hepatitis A virus grown in human diploid lung (MRC-5) cells, a cell line widely used for producing vaccines for human use (31). Table 4 lists the preparatory steps for the HAV HM-175/FI-1 vaccine. The final vaccine contained no detectable live virus by serial passage in cell culture or intravenous administration to susceptible owl monkeys. Monkeys immunized with this vaccine were also protected from challenge infection (Table 5)

In the first volunteer study, 8 seronegative adults received three 1.0 ml doses of HM-175/FI-1 one month apart in a manner identical to the monkeys. The antibody levels of humans were much lower than monkeys, so a fourth dose of vaccine was given 6 months after the first (32). Some recipients

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

TABLE 5. HEPATITIS A HM-175/FI-1 VACCINE
PROTECTION OF OWL MONKEYS*

Vaccine	Monkey	Fecal Antigen	Elevated ALT	Hepatitis A Antibody SPRIA **		Neutralizing	
				Pre	Post	Pre	Post
Vaccine	A18	No	No	89	93	2560	2560
	A19	No	No	74	93	2560	2560
	A34	No	No	76	89	160	640
	A44	No	No	82	92	160	40
	A48	No	No	84	94	160	640
Control	A38	Yes	Yes	–	95	<10	10240
	A43	Yes	Yes	–	93	<10	10240
	8102	Yes	Yes	–	81	<10	160

*Monkeys challenged with 10^6 infectious doses 7 weeks following 3rd dose of vaccine.
** Percent inhibition by HAVAB Test, Abbott Lab. (>50% is positive)

experienced slight transient stinging at the site of inoculation attributable to the formalin, but there were no systemic complaints. The peak GMT neutralizing antibody was 320 following the fourth dose, matching the level in the monkeys protected from challenge (Figure 1). Subsequent followup of the first 8 recipients has shown they all continue to have detectable neutralizing antibody titers of ≥10 for more than 3 years, significantly exceeding the measurable antibody titers to IG (Unpublished observations). Of equal importance, the serum antibody reacts with heterologous strains of human HAV obtained from five geographic regions (Table 6). It is expected that immunization with the HM-175 strain of hepatitis A should provide protection against all hepatitis A viruses.

A subsequent study in 42 soldiers at Fort Lewis, WA demonstrated that there was no important difference in antibody responses between the original 0,1,2,6 month dosage schedule and a 0,1,6 month schedule (33). It is probable that an inactivated hepatitis A vaccine can be given simultaneously with the 0,1,6 month dosage schedule of hepatitis B vaccine. This hypothesis is being tested presently in soldiers at Fort Campbell, KY.

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

IMPROVED HEPATITIS A VACCINES

The WRAIR HAV HM-175/FI-1 vaccine induced serum antibody in only 86% of the recipients; it should immunize 95% to be an acceptable product. The potency is deficient because the hepatitis A antigen concentration is estimated to be only 2-17 ng/1.0 ml dose, a remarkably low dose compared to licensed vaccines for other diseases. Furthermore, no adjuvant was added to the vaccine because the viral antigen was not purified. Preliminary evidence from animals suggests that a purified antigen with added alum adjuvant will triple the antibody levels of antigen alone.

TABLE 6. HEPATITIS A HM-175/FI-1 VACCINE
ANTIBODY REACTIVITY TO HEPATITIS A VIRUSES

Geographic Region	Virus Strain	GMT*
Australia	HM-175**	269
Kansas	LV-374	319
Alaska	FR AL	269
Germany	GR-8	269
Panama	PA-21	269
N. Africa	MBB	113

*Geometric mean neutralizing antibody titers of human sera
**Parent virus for vaccine

In 1989, large scale production of purified HAV began under USAMRDC contract following the recommendations of Dr. Dubois et al (34). Initially, 300 liters of cell culture harvest are being purified, inactivated with formalin, and adsorbed to alum or another potent adjuvant to yield an estimated 30,000 doses of Hepatitis A HM-175/FI-2 vaccine for initial testing during 1990. This vaccine will initiate a limited reserve for immunizing military personnel if commercial sources are not available in the near future.

Meanwhile, WRAIR investigators will evaluate commercially developed hepatitis A vaccines as they become available. Since the initial success of the HM-175/FI-1 vaccine, two major vaccine manufacturers have prepared inactivated whole virus vaccines of their own. The authors have already conducted preliminary safety testing of two products from Smith Kline Biologics, Inc. The most acceptable preparation, also derived

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE
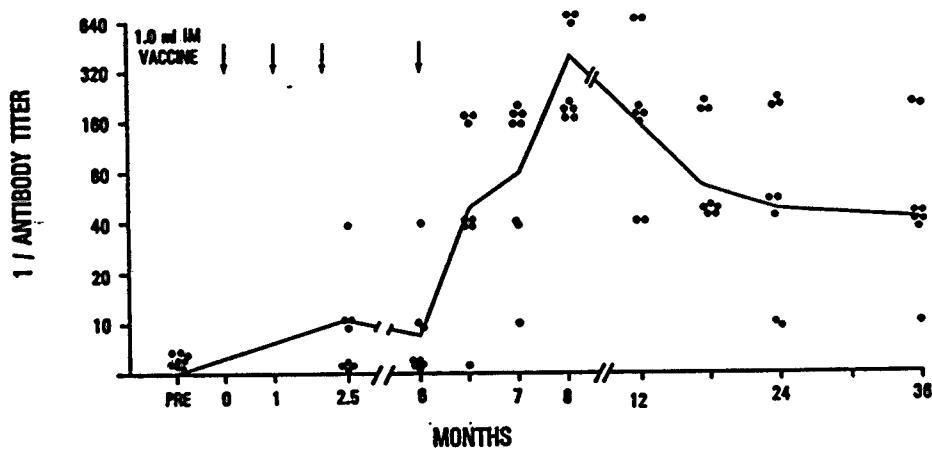
Figure 1. Duration of serum neutralizing antibody to WRAIR HAV
HM-175/FI-1 vaccine in 8 recipients.
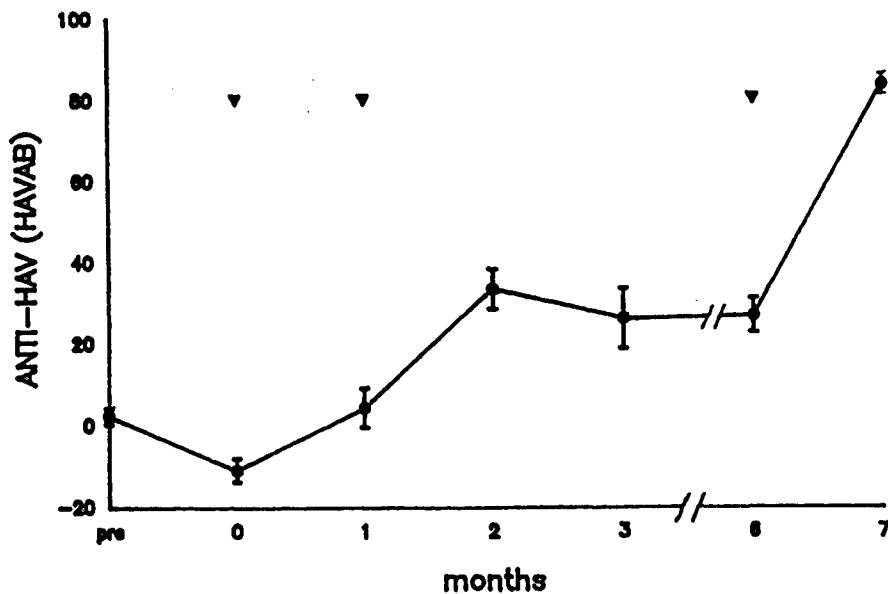
Figure 2. Serum HAV antibody levels (% inhibition by radio-
immunoassay) of 7 responders to the Smith Kline HM-175 vaccine.

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

from a HM-175 strain of hepatitis A cloned at the WRAIR, has proven to be entirely safe and to induce serum antibody following 3 doses (Figure 2). This product is currently being tested in 297 soldiers at Fort Campbell, KY to determine if the vaccine can be given simultaneously with hepatitis B vaccine and if hepatitis A vaccine can also be given safely by mechanical jet injection. The results of the Fort Campbell study should provide useful information on the feasibility of immunizing large numbers of soldiers assigned to deployment ready units.

VACCINE AVAILABILITY AND USE

It appears that the major manufacturing difficulties have been overcome. Manufacturers are initiating vaccine studies in children and adults. These studies will expand the experience on safety and immunogenicity and allow precise determination of optimal dosage schedules. The most difficult step will be to prove protective efficacy from natural infection. With a long incubation infection, efficacy studies may require years to complete. Nevertheless, efficacy trials are expected to begin in endemic areas in the next two years. Depending upon a successful outcome, US licensure of an inactivated hepatitis A vaccine may be possible within 5-10 years.

For Americans, the beneficiaries of a hepatitis A vaccine will include travelers to endemic areas, health care workers, child care center employees, food service workers, prisoners and all military recruits. At last, there is hope of preventing one of the major historic threats to military operations, acute viral hepatitis, type A.

REFERENCES

1. Bancroft WH and Lemon SM. (1984) Hepatitis A from the Military Perspective. In: Hepatitis A (RJ Gerety Ed) Academic Press, pp81-100.

2. Stokes J Jr and Neefe JR (1945) Prevention and Attenuation of Infectious Hepatitis by Gamma Globulin. J Amer Med Assoc 127:144-145.

3. Neel S. (1973) Medical Support of the U.S. Army in Vietnam 1965-1970. Department of the Army. Washington DC pp32-48.

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

4. Scott RM, Schneider RJ, Snitbhan R. and Karwacki JJ. (1981) Factors Relating to Transmission of Viral Hepatitis in a United States Military Population Stationed in Thailand. Am J Epidemiol 113:520-528.

5. Lemon SM, Lednar WM, Bancroft WH., Cannon HG, Benenson M, Park JH, Churchill FE, Tezak RW, Erdtmann FJ, Kirchdoerfer RG, Lewis PG, James JJ and Miller RN. (1982) Etiology of Viral Hepatitis in American Soldiers. Am J. Epidemiol. 116:438-450.

6. Bennion SD, Burton FV, Barrett JR and Gocke DT. 1987. A Prospective Study of the Etiology of Hepatitis in the U.S. Army Wurzburg Community from 1978 to 1979. Milit Med 152:80-82.

7. Aronson, NE and Palmer BF. (1988) Acute Viral hepatitis in American Soldiers in Korea. South Med J 81:949-951.

8. Dienstag JL, Routenberg JA, Purcell RH, Hooper RR and Harrison WO (1975). Ann Int Med 83:647-650.

9. Benenson MW, Takafuji ET, Bancroft WH, Lemon SM, Callahan MC and Leach DA (1980) A Military Community Outbreak of Hepatitis Type A Related to Transmission in a Child Care Facility. Am J Epidemiol 112:471-481.

10. Hepatitis Surveillance Report No. 52. Atlanta: Centers for Disease Control, 1989. pp1-32.

11. Lednar WM, Lemon SM, Kirkpatrick, JW, Redfield RR, Fields ML and Kelley PW. (1985) Frequency of Illness Associated with Epidemic Hepatitis A Virus Infections in Adults Am J Epidemiol 122: 226-233.

12. Lemon SM. (1985) Type A Viral Hepatitis. New Eng J Med 313:1059-1067.

13. Shimizu YK, Shikata T, Beninger PR et al. (1982) Detection of Hepatitis A Antigen in Human Liver. Infect Immun 36:320-324.

14. Lemon SM, Brown CD, Brooks DS, Simms TE and Bancroft WH. (1980) Serum Immunoglobulin M Response to Hepatitis A Virus Determined by Solid-Phase Radioimmunoassay. Infect Immun 28:927-936.

15. Kurane I, Binn LN, Bancroft WH and Ennis FA. (1985) Human Lymphocyte Responses to Hepatitis A Virus-Infected Cells:

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

Interferon Production and Lysis of Infected Cells. J Immunol 135:2140-2144.

16. Sjogren MH, Tanno H, Fay O, Sileoni S, Cohen BD, Burke DS and Feighny RJ. (1987) Hepatitis A Virus in Stool During Clinical Relapse. Ann Int Med 106:221-226.

17. Centers for Disease Control. Protection Against Viral Hepatitis: Recommendations for the Immunization Practices Advisory Committee (ACIP) MMWR 1990;39 (No. RR-2):1-26.

18. Ticehurst J, Cohen JI and Purcell RH (1988) Analysis of Molecular Sequences Demonstrates that Hepatitis A Virus is a Unique Picornavirus. In Zuckerman AJ (ed): Viral Hepatitis and Liver Disease. New York: Alan R Liss Inc. pp33-35.

19. Provost PJ and Hilleman MR (1978) An Inactivated Hepatitis A Virus Vaccine Prepared from Infected Marmoset Liver. Proc Soc Exp Biol Med 159:201-203.

20. Provost PJ, Banker FS, Giesa PA, McAleer WJ, Buynak EB and Hilleman MR. (1982) Progress Toward a Live, Attenuated Human Hepatitis A Vaccine. Proc Soc Exp Biol Med 170:8-14.

21. Emini EA, Hughes JV, Perlow DS and Boger J. (1985) Induction of Hepatitis A Virus-Neutralizing Antibody by a Virus-Specific Synthetic Peptide. J Virol 55:836-839.

22. Provost PJ, Bishop RP, Gerety RJ, Hilleman MR, McAleer WJ, Scolnick EM and Stevens CE. (1986) New Findings in Live, Attenuated Hepatitis A Vaccine Development. J Med Virol 20:165-175.

23. Binn LN, Lemon SM, Marchwicki RH, Redfield RR, Gates NL and Bancroft WH. (1984) Primary Isolation and Serial Passage of Hepatitis A Virus Strains in Primate Cell Cultures. J Clin Microbiol 20:28-33.

24. Lemon SM, Binn LN and Marchwicki RH. (1983) Radioimmunofocus Assay for Quantitation of Hepatitis A Virus in Cell Cultures. J Clin Microbiol 17:834-839.

25. Lemon SM and Binn LN. (1983) Serum Neutralizing Antibody Response to Hepatitis A Virus. J Infect Dis 148:1033-1039.

BANCROFT, BINN, ECKELS, LEDUC, SJOGREN, DUBOIS, MARCHWICKI, HOKE

26. Stapleton JT, Jansen R and Lemon SM. (1985) Neutralizing Antibody to Hepatitis A Virus in Immune Serum Globulin and in the Sera of Human Recipients of Immune Serum Globulin. Gastroenterology 89:637-642.

27. LeDuc JW, Escajadillo A and Lemon SM. (1981) Hepatitis A Virus Among Captive Panamanian Owl Monkeys. Lancet 2:1428.

28. LeDuc JW, Lemon SM, Keenan CM, Graham RR, Marchwicki RH and Binn LN. (1983) Experimental Infection of the New World Owl Monkey (Aotus trivirgatus) with Hepatitis A Virus. Infect Immun 40:766-772.

29. Trahan CJ, LeDuc JW, Staley EC, Binn LN, Marchwicki RH, Lemon SM, Keenan CM and Bancroft WH. (1987) Induced Oral Infection of the Owl Monkey (Aotus trivirgatus) with hepatitis A Virus. Lab Animal Sci 37:45-50.

30. Binn LN, Bancroft WH, Lemon SM, Marchwicki RH, LeDuc JW, Trahan CJ, Staley EC, and Keenan CM. (1986) Preparation of a Prototype Inactivated Hepatitis A Virus Vaccine from Infected Cell Cultures. J Infect Dis 153:749-756.

31. Binn LN, Bancroft WH, Eckels, KH, Marchwicki RH, Dubois DR, Asher LVS, LeDuc JW, Trahan CJ, and Burke DS. (1988) Inactivated Hepatitis A Virus Vaccine Produced in Human Diploid MRC-5 Cells. In Zuckerman AJ (ed): Viral Hepatitis and Liver Disease. New York: Alan R. Liss Inc. pp91-93.

32. Sjogren MH, Eckels KH, Binn LN, Dubois DR, Hoke CH, Burke DS and Bancroft WH. (1988) Safety and Immunogenicity of an Inactivated Hepatitis A Vaccine. In Zuckerman AJ (ed): Viral Hepatitis and Liver Disease. New York: Alan R Liss Inc. pp94-96.

33. Sjogren MH, Hoke CH. Binn LN, Eckels KH, Dubois DR, Lyde L, Tsuchida A, Oaks S Jr, Marchwicki R, Lednar W, Chloupek R, Ticehurst J, and Bancroft WH. (1990) Safety and Immunogenicity of an Inactivated Hepatitis A Vaccine. Evaluation of Two Dosage Schedules. (Submitted).

34. Dubois DR, Eckels KH, Ticehurst J, Binn LN, Timchak RL, Barvir DA, Rankin CT and O'Neill SP. (1990) Large-Scale Purification of Inactivated Hepatitis A Virus by Centrifugation in Non-Ionic Gradients. (Submitted).

HIERARCHICAL ROUTE PLANNER (U)

Mr. John R. Benton*
Ms. Anne Brink
Artificial Intelligence Division
U.S. Army Engineer Topographic Laboratories (USAETL)
Fort Belvoir, Virginia 22060-5546

1. INTRODUCTION

Anyone who has struggled to find the most direct route between two locations on a city map where there is no single road or even just two or three roads directly connecting the two locations is familiar with the general problem of planning an optimum path through a complicated tangle of roads. In planning a route through a city one must consider not only distance but also the number of traffic lights and quantity of traffic at a given time of day. A battlefield commander may not have to worry about traffic lights when planning a route, but he must consider positions of enemy and friendly forces, lines of sight from enemy observation points, availability of places of concealment from aerial observation, choke points where his forces could be ambushed, and weather conditions that could affect the mobility of his vehicles.

Currently, there is no capability of automated route finding available for operational use in the Army. Such a capability would be particularly useful in situations that require a rapid reaction to counter an enemy threat or changing conditions. One example of a possible Army application is the requirement that a friendly force's commander be able to predict the enemy's objective and the routes the enemy will use to advance toward their objective. Military doctrine often calls for offensive forces to advance in three separate columns. The defending forces must use their knowledge of the doctrine and tactics of the opposing force as well as knowledge of the terrain to anticipate and possibly to counteract the movement of their enemy. The computer should therefore compute three optimum noncompeting paths as well as suboptimum alternatives. The three routes must be noncompetitive (i.e. have no road segments in common) since there would be traffic jams if two separate columns had to share the same road.

Terrain analysis performed in support of battlefield planning is tedious and time consuming. A detailed terrain analysis of a strategically important area may take as long as a month to perform. The Army's Digital Topographic Support System (DTSS), currently under development at USAETL, does not automate terrain analysis. Instead, it provides computer support such as computation of lines-of-sight and replaces the traditional mylar map overlays with an automated Geographic Information System. The analyst must still use his intuition in determining military corridors and key terrain.

Since accurate knowledge of the battlefield is critical to successful operations and it is not feasible to analyze in advance all potentially important areas, automated terrain reasoning (and route planning

in particular) provides a possible way to improve the productivity of terrain analysts and permit a more rapid response to changing conditions.

In developing a computer model for route planning over a large area, a prime concern must be to avoid what is known as "combinatorial explosion." For example, a four-levels-deep full binary tree has 15 branch points or nodes while a 20-levels-deep full binary tree has more than a million nodes. If the map data consists of a grid at 10 meter spacings, an unrestricted search algorithm would generate a search tree with literally millions of nodes[*] in order to plan a route for a distance of less than 200 meters. Obviously, no one would actually use such a simple-minded technique, but when route planning over a distance of several kilometers is required, the almost inevitable result of planning a route at the pixel level is combinatorial explosion. An alternative approach is to reduce the size of the search space before the route planning algorithm is invoked.

In an earlier paper,[1] one of the authors developed the concept of converting a Cross Country Mobility Map (CCM) to a binary go/no-go map by setting a velocity threshold such that terrain on which a tank can sustain some minimum velocity will be marked as *go* and all other areas will be
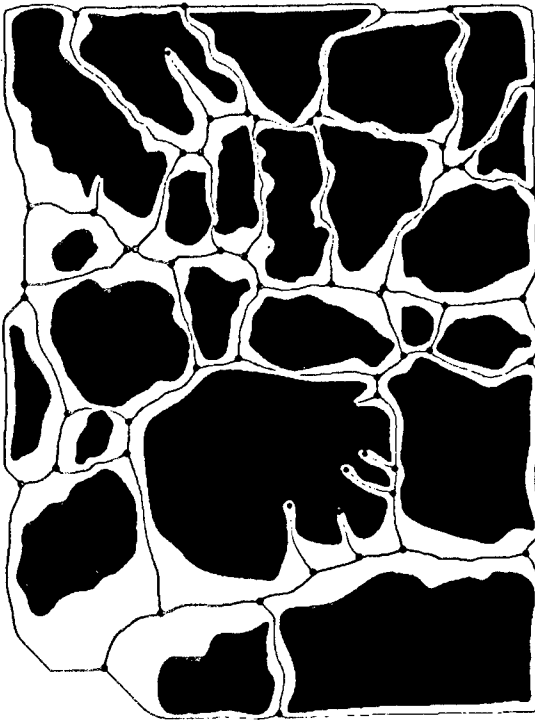
Figure 1. Cross-Country-Mobility Map with skeleton of "GO" areas superimposed. Black represents obstacles.
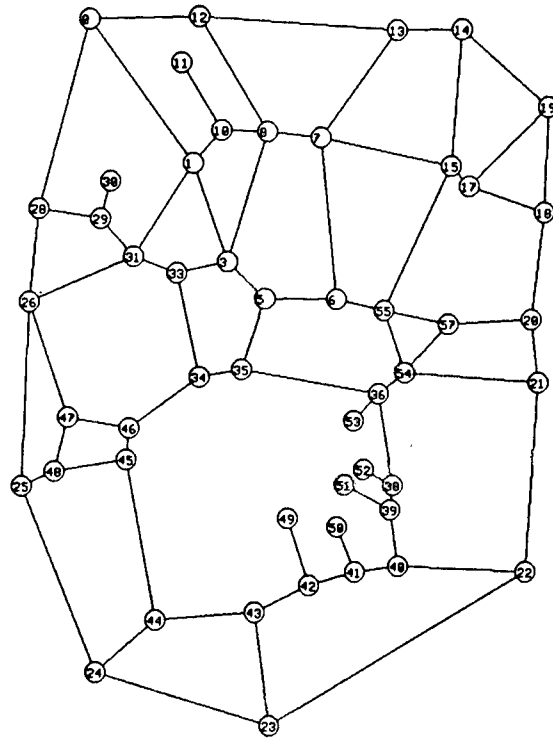
Figure 2. Graph Representation of Mobility Map after Merger of Closely Adjacent Points

---

[*] A node indicates a point along a path where there is a choice in which direction to proceed.

marked as *no-go*. A line-thinning algorithm is then used to generate the skeleton of the binary map. Branches in the skeletal structure are caused by obstacles that force a vehicle to fork either to the left or to the right in order to go around the obstacle. A vectorizing routine traces all lines, finds the nodes connecting lines and generates a graph-theoretic description of the skeleton. A tree-searching algorithm can then be used to find optimum paths across the graph. Since the skeleton is simply the center line of the *go* areas, the skeleton, in general does not provide the exact optimum path between two nodes. A pixel-level route planner that searches all adjacent pixels of the mobility map can be used to compute the precise path between two nodes. This pixel-level planner is useful only over short distances since there will be an exponential increase in the number of adjacent pixels to be explored as the path length increases. Thus, the solution is to use a pixel-level route planner to generate the precise route between nodes and a graph level route planner to efficiently plan a route over a larger area but without detail about the precise path that was followed. The traversal time between nodes computed by the pixel-level route planner will be used as weights by the graph-level route planner. The graph-level planner uses the traversal times generated by the pixel-level planner. The resulting system is a two-level hierarchical route planner. This is the approach that was selected for use on this project. Intelligent search methods can be applied at both route planner levels. Figure 1 is a mobility map of synthetic data drawn by a terrain analyst. The black areas are obstacles and the white areas have been thinned to a skeleton. The dots positioned at the ends of lines and at intersections indicate where the vectorizing routine located nodes in the graph. In Figure 2, closely adjacent nodes have been merged with straight lines drawn between the nodes. For example, nodes 15 and 17 of Figure 2 correspond to three nodes in Figure 1. Figure 3 is a block diagram of the overall Hierarchical Route Planner.

Geographic Information System. This project is using the QUILT Geographic Information System.[2] It is quad-tree[3] based, combined with an efficient storage mechanism for retrieving very large
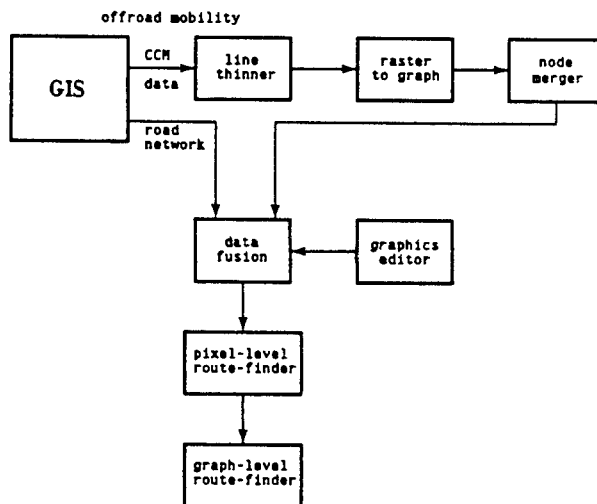
Figure 3. Hierarchical Route Planner data flow diagram.

BENTON, BRINK

sets of data. QUILT contains very efficient mechanisms for the rapid computation of intersection and union operations of two map overlays. QUILT provides the capability to convert mobility maps into the raster format required by the line thinner module. QUILT was installed on a Sun Workstation using the Version 11 X Windows[4] package.

Line Thinner. A line thinning algorithm developed by Chang and Suen[5] was implemented in Common Lisp on a Lisp machine. This algorithm was designed so that all scan lines can be thinned simultaneously on a parallel machine. An artifact of this algorithm is that lines with slope of minus one are connected only at the diagonal corners (eight-neighbor connectivity) while lines with slope of plus one have adjacent pixels that are connected on the sides (four-neighbor connectivity). The original mobility map may also have adjoining regions connected only by the corner of one pixel touching the corner of a pixel of the adjoining region (eight-neighbor connectivity) and this connectivity must be preserved in the skeleton. A 500 by 700 pixel raster currently requires 15 minutes to run on a Symbolics 3600 computer.

Raster-to-Graph. This module scans the screen array until it finds the first nonzero pixel. It then looks at the eight neighbors of the pixel and traces along the line until it finds a pixel that has two neighbors which are not adjacent. This pixel becomes the first node, and its value is changed to some other nonzero value so that it will have a different color on the display and more importantly be invisible to the line following algorithm. A Chain code, consisting of the vector directions from one pixel along the path to the next, is used to store the paths, with values from zero to seven used to represent the eight possible directions. As the line follower advances along the line, each pixel has its color changed so that the line follower will not circle back on itself. This path follower correctly located the 60 nodes of the graph in Figure 3. All pixels of the graph were traversed.

Node Merger. The line thinner often generates adjacent nodes which need to be merged into a single node. This module scans all detected arcs and eliminates closely-spaced nodes. When four or more paths of the CCM map converge to a single junction, the area of the intersection is frequently quite large compared to the width of the paths. The node-merger module merges nodes if the separation of the nodes is less than a threshold value. This value is currently a constant which is transferred to the node-merger function on the argument list. The threshold value used in generating Figure 2 was 25 pixels in vertical and horizontal directions. A total of 10 nodes were merged with adjacent nodes. A future enhancement will be to have the module check to see if both nodes actually are located at the same junction of the cross-country-mobility map.

Graphics Editor. This module provides a capability for allowing the user to modify the generated graph. Changes may be necessary due to new information obtained from aerial photographs, reconnaissance, or other data available to the user. The user interface of this module has not been completed but most of the functional capabilities required by this module have been coded. One submodule allows the user to add nodes and arcs to the graph by tracing the arc-paths with a mouse pointer. Spline curves are fitted to the traced lines and the data structures required by the graph-level route planner are generated. When a path is being traced, if there is a sharp turn in the traced path, then a separate spline curve will be used on each side of the turn. A discontinuity in slope can thereby be accommodated. The route planner can be run using data generated solely by the mouse-tracker. Functions contained in the node-merger module will also be used in the graphics editor.

Data Fusion. When there is both road data in vector format from the QUILT GIS and a graph derived from the CCM, a capability to merge road data will be required. Development of the data-fusion module will not begin until the GIS has been integrated into the route planner. Functionally, this module will be similar to the raster-to-graph module.

Pixel-Level Route Planner. Several pixel-level route planners have been reported in the literature. Denton and Froeberg[6] used dynamic programming to select a route across a composite cost surface on which the actual slope of the terrain at any given point is increased by an amount proportional to the localized threat of being at that location. The route planner tends to avoid areas of high slope and thus an aversion to threat areas is introduced. The weighting factors for the threats are determined by an expert system. The Route Planner Development Workstation[7] includes a number of pixel-level route planners that use a cost surface conceptually similar to that of Denton and Froeberg. The RPDW software package was developed for USAETL through an agreement with NASA. Other pixel-level route planners have been developed by Hughes Aerospace[8] and U.S. Army Waterways Experiment Station[9] A pixel-level route planner similar to one of these route planners will be used to compute the weights used in the graph-level route planner and to determine the exact path between nodes.

Graph-Level Route Planner. An algorithm essentially equivalent to the A* algorithm developed by Hart, Nilsson, and Raphael[10] was used to search the graph structure for an ordered listing of noncompetitive optimum routes between two given nodes. Each route is represented by a linked list of nodes, with the first node being the start and the final node being the destination node. The route with the lowest cost is listed first. Since the routes are noncompetitive, no two routes can share a path and thus they cannot share two successive nodes in their respective paths. The program is currently more restrictive in that no two paths can share a road intersection. Depending on the road widths at intersections, this may or may not be a necessary restriction.

The major difference of the Multiple Route Finder[11] (MRF) algorithm used here from A* is that A* halts as soon as the first optimum route is discovered, while MRF keeps searching until either there are no more unexplored paths or the required number of routes has been found. A* keeps two separate lists of nodes called OPEN and CLOSED. OPEN is the list of nodes that have not yet been explored. Initially, it contains only the start node. When a node is explored, it is moved from OPEN to CLOSED, and its descendants are put on the OPEN list. However, if one of these nodes has previously been reached by another route, then the more costly of the two routes must be pruned, and only the cheaper of the two routes will be retained. The MRF algorithm differs from A* in that the OPEN and CLOSED list are kept as a single linked list, with a slot at the node address to specify whether the node type is TERMINAL (CLOSED), NONTERMINAL (OPEN), PRUNED, or DESTINATION. A* terminates when the destination node is reached, in contrast to MRF, which continues to search for additional routes; therefore an explicit DESTINATION type is needed.

The A* algorithm uses a heuristic function to estimate the cost to go from the current node to the destination node. This function is usually called $h(n)$, where n is the current node. The total estimated cost associated with the node n is $f(n) = g(n) + h(n,)$ where $g(n)$ is the actual cost in traveling from the start node to the node n. Thus, $f(n)$ is the estimated total cost for traversing a path from the start node to the point n and then on to the destination node. A heuristic function can be shown to be "admissible" if it is monotonically nondecreasing and nonnegative. If the heuristic function is admissible, then the algorithm is guaranteed to terminate with the optimum solution. Conceptually, the simplest heuristic that meets these requirement is the euclidean distance from the node n to the destination node. The actual cost can equal but never be less than this straight line distance from n to the destination. However, the traversal time will normally be used as the weight rather than the actual distance. The cost to completion time must then be computed on the basis of the highest velocity class of the mobility map.

BENTON, BRINK

2. Investigation

Figure 4 is an actual mobility map made using the Condensed Army Mobility Model System (CAMMS).[12] The real world is much more complex than the synthesized map of Figure 3. Each pixel in Figure 4 corresponds to approximately 100 meters. In Figure 4, there are isolated islands or blips of unrestricted mobility in the middle of no-go areas, blips of no-go in the the midst of unrestricted mobility and between these two extremes there is every possible juxtaposition of velocity ranges. Figure 5 shows the the result of thresholding the three highest velocity ranges as "go" with the line-thinned skeleton superimposed. Each of the blips of Figure 4 results in the creation of additional nodes and the resulting skeleton is so complex that it is practically useless. Since the blips are responsible for the proliferation of nodes, the solution is to remove all isolated blips. First a threshold must be established for the largest blip that is to be removed. Next, criteria must be established to determine if the blip is truly isolated or is connected to an adjoining region. A given pixel has either four neighbors or eight neighbors depending on whether adjacency is defined in terms of the four principal direction or the four principal directions plus the diagonal directions. A blip is considered isolated unless it has four-neighbor connectivity to another region. The alternative would have been to require a region to not even have eight-neighbor connectivity in order to be considered isolated. This more stringent restricted would have resulted in many fewer blips being eliminated. An algorithm was previously developed for removing isolated local minima from drainage network data[13]. This algorithm was adapted to remove both the islands of *ones* from the ocean of *zeroes* and the lakes of *zeroes* from the continents of *ones*. Figure 6 shows the result of smoothing all blips of Figure 5 that contain fewer than 100 pixels. In comparing the circled area in the upper left corner of Figure 6 to the corresponding area of Figure 5, a loss of connectivity as result of smoothing is observed. A blip which was removed bridged a gap by having diagonal connectivity at each end of the blip. This effect is a consequence of using eight-connectivity for the skeleton but four-connectivity for smoothing. If eight connectivity had been used for smoothing, the resulting graph would have been much more complex and difficult to use. Methods for dealing with this problem will be discussed later in this paper.

Figure 7 shows the results of vectorizing the skeleton of Figure 6 with the criteria that nodes with a separation of no more than eight pixels would be merged together. Also, short dangling limbs were eliminated unless the removal of the corresponding nodes would cause a triangle to degenerate into two points with two alternative distances separating the points. Figure 8 shows the results of requesting three independent routes between nodes 38 and 67. The shortest path is indicated by the heavy line and had an associated weight of 209, the second best route is indicated by the cross hatching and had a weight of 295 and the third best route is indicated by line with superimposed dots. Its weight is 366. It must again be emphasized that this graph can not simply be superimposed on a map sheet with the expectation that one can traverse along straight lines between the nodes. The actual procedure will be to position the nodes on the map, repositioning any nodes that happen to fall on a *no-go* blip that was removed in the smoothing process. The node must be repositioned into an adjacent *go* area. A pixel-level route planner must then be invoked to compute the route between the nodes along the path. This path is still not optimized since the position of the node points can not be expected to fall on top of the optimum path. Relaxation methods can be used to make the final adjustment of the total path between the start node and the end node. As a simplified example, the path of 38-88-86 in Figure 8 can be determined as follows. The pixel-level route planner computes the optimum path between the nodes 38 and 88 and between the nodes 88 and 86. A point half way along the computed path of 38 to 88 is defined as a new node. Similarly, a point half way along the computed path between 88 to 86 is defined as a second new node. The pixel-level route planner then computes the path between the two newly defined nodes.

BENTON, BRINK

Figure 5 Thresholded Mobility Map with Superimposed skeleton

Unrestricted Velocity

Restricted Velocity

Slow Velocity

Very Slow Velocity

Impassable

Urban
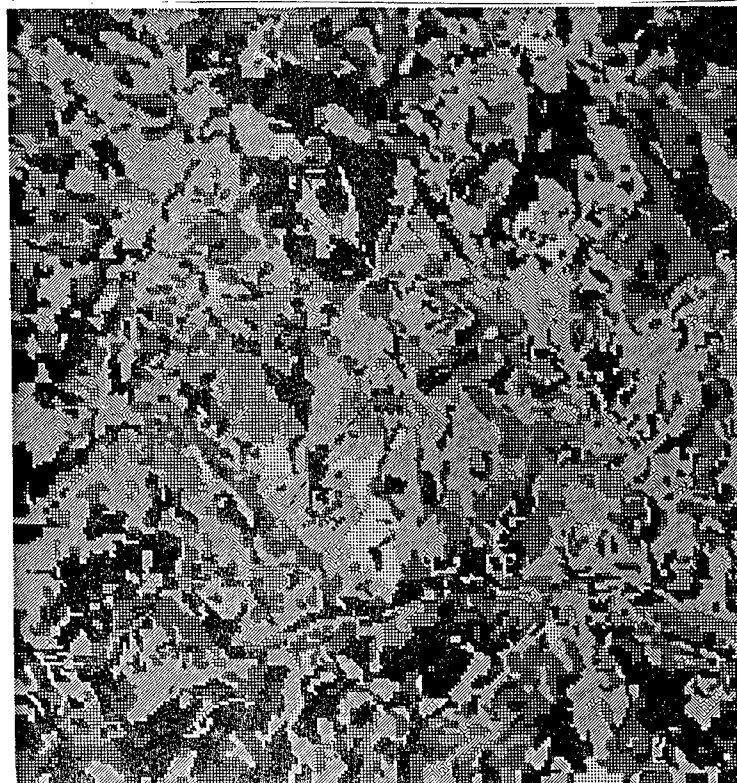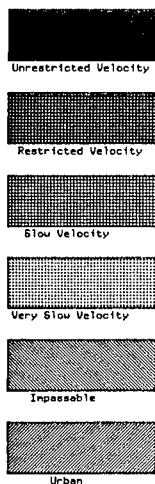
Figure 4 Mobility Map derived from CAMMSDATA
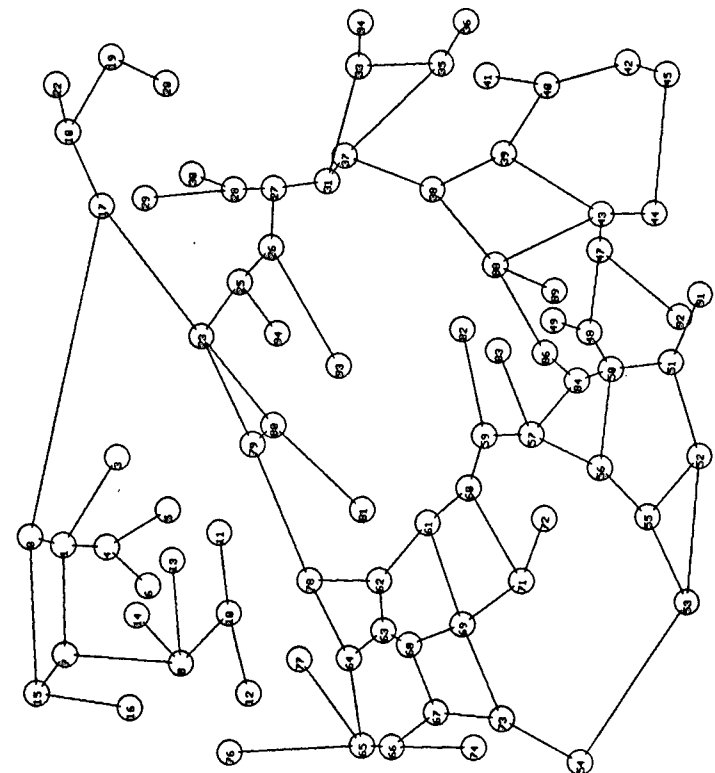
Figure 7.
Graph Representation of Mobility
Map of figure 6.
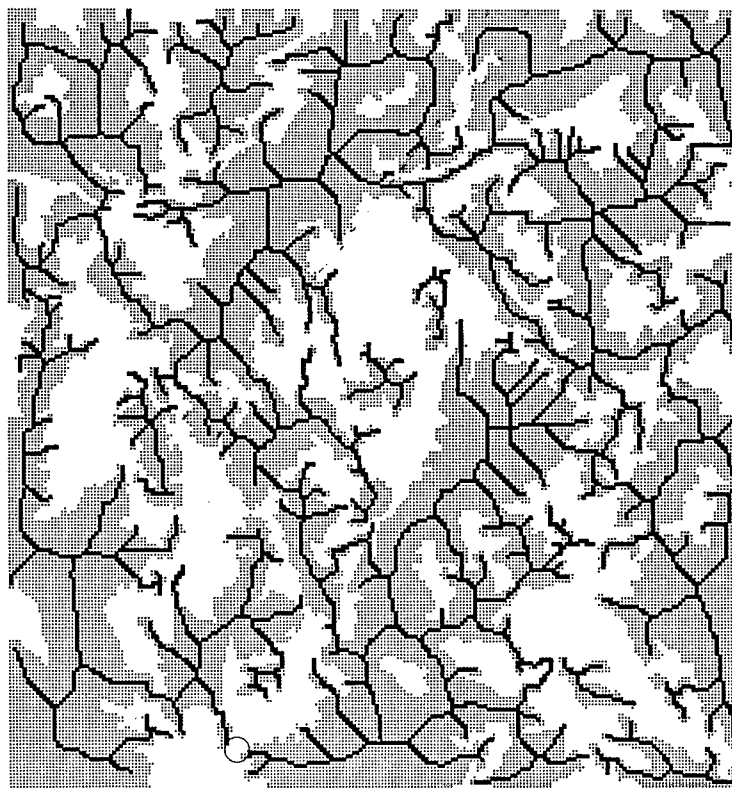
Figure 6.
Thresholded Mobility Map after
smoothing and with Superimposed
skeleton.
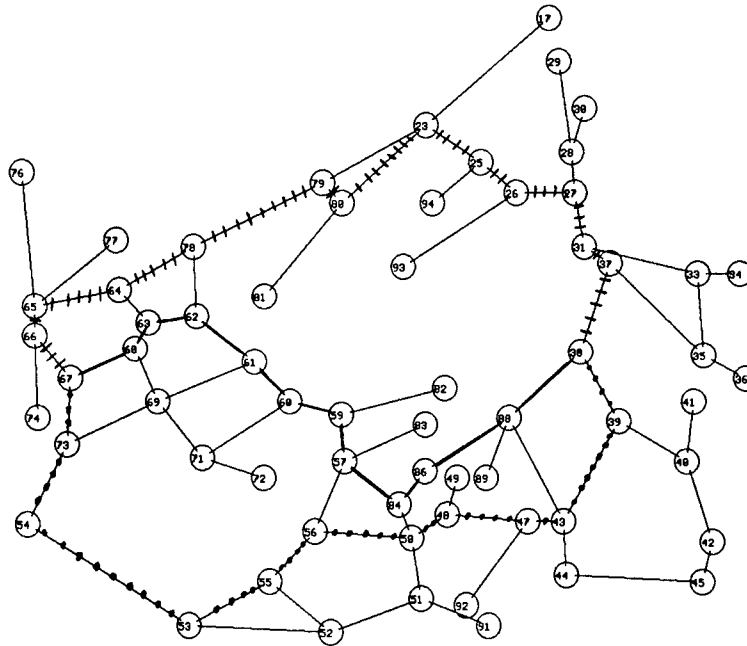
Figure 8. Three optimum Routes from Node 38 to Node 67.

One important property of the line thinning algorithm has been ignored up to now. The number of iterations of the algorithm required to generate each pixel of the skeleton can be saved and stored as part of the skeleton. If F(x,y) is the skeleton and g(x,y) contains the corresponding number of iterations used to produce the skeleton then for some x',y' located on the skeleton, g(x',y') is a direct measure of the width of the mobility corridor at the point x',y'. If for a given path between two nodes, all values of g(x,y) along that path exceed some minimum constriction then that path is suitable as a military corridor. Furthermore, the point at which g(x,y) is a minimum along that path is a potential chokepoint. This modification of the line thinner will be implemented in the near future.

An expert system can be useful both in managing the system and in application of military doctrine to route planning. The first important decision is the selection of an expert system shell. The requirement is not for a shell which can efficiently handle hundreds or thousands of rules but instead for a simple system in which typically a rule set will have no more than 20 rules. A complex algorithm to determine which rule will fire next is not needed - all of the rules must be applied to the data. Another critical consideration is the ability to have close interaction between the expert system shell and LISP. The rules must be able to incorporate calls to LISP functions. Finally, there is a preference that the expert system shell support the Common LISP Object System[14] (CLOS). However, since the object-oriented CLOS is still very new, expert system shells incorporating CLOS are not currently available. Each shell tends to use its own proprietary object oriented language. We decided to modify the expert system shell in Winston and Horn[14] to include the capability to evaluate LISP expressions. Their expert

BENTON, BRINK

system shell includes a pattern matcher and instantiation of variables. An example of a rule in Winston and Horn is as follows:

```
(species-rule        (?animal is a ?species)
                     (?animal is a parent of ?child)
                     (?child is a ?species))
```

where the last clause is the *then* part of the rule. The syntax of Winston and Horn uses (? animal) to indicate a variable. The shell was modified to use the less cumbersome ?animal. Finally, the syntax was modified so that the expression (evalu (contains ?route-segment 'chokepoint)) would instantiate ?route-segment to a route-segment contained in the data base, evaluate the LISP expression and if the rule fires, then the rules assertion can be added to the data base. The LISP function contains could also store the fact that the route segment contains a chokepoint. Several examples of how an expert system can be used include the following:

(1) Each of the *go* blip removed during the smoothing process can be examined to see if it has two separate points with diagonal connectivity to the surrounding *go* areas. If so, will its reincorporation into the graph have a significant effect on the graph. In addition, if the minimum allowable constriction is two hundred meters, then two diagonnally connected pixels each 100 meters square obviously do not meet the 200 meter minimum constriction requirement and the effect of the blips on connectivity need not be considered.

(2) Denton and Froeberg[6] gives an example of a rule that could be used in their expert system:

if   the current best route candidate goes through a choke point, and

if   recent enemy order of battle intelligence suggests that the choke point may be under enemy control before it can be reached and

if   coordinated fire support will not be available to impede the enemy advance toward the choke point,

then consider the performance of alternative , less direct (detour) route candidates associated with an altered composite map that penalizes use of the current choke point.

Where Denton and Froeberg would modify the composite map, our approach would be to remove the arc of the graph that contains the dangerous choke point and recompute the route. This recomputed route would then be compared to the previously computed route.

A final application for an expert system is to find locations where it might be possible to breech a narrow obstacle in order to effect surprise on the opposing force or to significantly reduce the traversal time. Skeletons of the obstacles can also be computed and superimposed on the thresholded mobility to show not only where the *go* areas are but also where there are obstacles that totally prevent mobility.

Conclusions

The Hierarchical Route Planner with the extensions for incorporation of an expert system provides the nucleus of an Automated Terrain Reasoning System. The two-level planning provides capability to do planning over much longer distances than is possible with pixel level route planner. Possible applications for route planning include minefield site prediction, siting of air defense artillery weapons, autonomous vehicles and route selection for tanks and other off-road vehicles.

BENTON, BRINK

3. REFERENCES

1. J. R. Benton "Hierarchical Route Planner, *Applications of Artificial Intelligence VI*, SPIE **937** pp. 428-433, (1988).

2. C. A. Shaffer, H. Samet, and R. Nelson, *QUILT, A Geographic Information System Based on Quadtrees*, Center for Automation Research, University of Maryland, College Park Maryland, 20742, Report No. CAR-TR-307 CS-TR-1885, (1987).

3. H. Samet, "The Quadtree and Related Hierarchical Data Structures,"*ACM Computing Surveys*, **16** *2*, 187-260 (1984).

4. R. W. Schiefler and J. Gettys, "The X-Window System," *ACM Transactions on Graphics*, **5** *2*, 79-109 (1986).

5. T. Y. Zhang, and C. Y. Suen, "A Fast Parallel Algorithm for Thinning Digital Patterns," *Comm. of the ACM*, **27** *3*, 236-239 (1984).

6. R. V. Denton and P. L. Froeberg, "Applications of Artificial Intelligence in Automated Route Planning," *Applications of Artificial Intelligence*, SPIE **485** pp. 126-133, (1984).

7. J. Cameron, B. Cooper, A. Mishkin, K. Holmes, *Route Planner Development Workstation - Functional Description*, Jet Propulsion Laboratory, Pasadena, California, Report JPL D-2733 Vol. 1, (1985).

8. Joseph Mitchell, David Peyton and David Kiersey, "Planning and Reasoning for Autonomous Vehicle Control," *International Journal of Intelligent Systems*, **2** 129-189 (1987).

9. G. B. McKinley, "Optimum Route Evaluation Using CAMMS," *U.S. Army Operations Research Symposium*, Vicksburg, MS (Proceedings in Press).

10. P. E. Hart, N. J. Nilsson, and B. Raphael, "A Formal Basis for the Heuristic Determination of Minimum Cost Paths", *IEEE Trans. System Science and Cybernetics*, **SSC-4** *2*, 100-107 (1968).

11. J. R. Benton, *Automated Route Finder for Multiple Tank Columns*, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia (1987).

12. T. C. Falls, et. al. *A Users Guide for CAMMS Version 2.0*, U.S. Army Waterways Experiment Station, Vicksburg, MS. 1989.

13. W. W. Seemuller, "The Extraction Of Ordered Vector Drainage Networks From Elevation Data", *Computer Vision, Graphics and Image Processing*, **47**, 45-58 (1989).

14. Patrick Henry Winston and Berthold K. P. Horn, *LISP*, 3rd Edition, Addison-Wesley Publishing Company, 1989.

Nonlinear Metal Composites for Guided Wave Devices in Optical Computing (U)

M. J. Bloemer, Dr., P. R. Ashley, Dr., J. W. Haus, Dr.†,
and C. R. Christensen, Dr.
Research Directorate, Research, Development, and Engineering Center, U.S. Army Missile
Command, Redstone Arsenal, AL
35898-5248

Introduction

Materials with third order optical nonlinearities have received significant attention in recent years. This interest has in part been motivated by the realization of "all-optical" devices [1]. An all-optical device uses light to control light through an intensity dependent index of refraction. The theoretical limit on switching times in all-optical devices is subpicosecond. Traditional integrated optics devices such as Mach-Zehnder interferometers and directional couplers rely on the electrooptic effect. The same devices can be made of materials with third order nonlinearities.

In 1987, Gabel et al [2] were the first to report degenerate four-wave mixing (DFWM) in planar waveguides with a nonlinear guiding medium. The waveguides were fabricated in semiconductor-doped glass (Corning 3-69) by an ion exchange process. In the present paper, we report DFWM in a planar waveguide consisting of gelatin doped with gold particles. Experiments by Hache et al [3] on small gold particles indicate a value of $\chi_m^{(3)} = 10^{-8}$ esu at $\lambda = 532$ nm with response times of < 5ps. We have obtained conjugate reflectivities near 1% for gold doped waveguides.

† Permanent Address: Physics Department
                      Rensselaer Polytechnic Institute
                      Troy, NY 12180-3590

Theory

Effective-medium theory has found uses in disparate fields of research, including particle diffusion, thermal conduction, elasticity and electromagnetism. Most of the results have been formulated for linear media, but the thrust has been toward nonlinear phenomena in recent years. In electromagnetism, this progress has been spurred by the experimental findings of large enhancements of the optical nonlinearities at surface plasmon resonances.

The theoretical calculations correlate linear absorption data along with the DFWM results. The structure and position of the surface plasmon resonance is determined by the dielectric properties of the conducting and host material and more sensitively, by the shape of the particle.[4,5] In this article, we restrict our discussion to spherical particles, since TEM analysis of the samples show the particles to be approximately spherical in shape.

The linear absorption in the medium is calculated from the complex, linear dielectric constant of the composite medium which in the small concentration limit is given by:

$$\bar{\epsilon}_L = \epsilon_h + f\gamma^m\left(\epsilon_{mL} - \epsilon_h\right) \tag{1}$$

where the enhancement factor γ^m is:

$$\gamma^m = \frac{3\,\epsilon_h}{2\,\epsilon_h + \epsilon_{mL}}, \tag{2}$$

and ϵ_h is the dielectric function of the host dielectric medium; ϵ_{mL} is the linear, complex dielectric function of the metal particles and f is the volume fraction of metal particles. The average dielectric function is complex in the medium, the linear absorption of the composite medium is related to the imaginary part of the effective dielectric function:

$$\propto = \frac{\omega^2 \operatorname{Im}(\bar{\epsilon}_L)}{c^2 k} = \frac{\omega^2}{c^2 k}\epsilon''_m\left|\gamma^m\right|^2. \tag{3}$$

where we have taken the metal dielectric function as: $\epsilon_{mL} = \epsilon'_m + i\,\epsilon''_m$ and we assume that the absorption in the host dielectric is negligible. The wave number of the medium

is directly related to the real part of the complex dielectric function $k = \omega^2\,Re\,(\bar{\epsilon}_L)/c^2$ where ω is the angular frequency of the light. The main point to note from Eq. (3) is that the absorption is proportional to the square of the amplitude of the enhancement factor. This will be used in the analysis of the experimental results discussed below.

In a composite dielectric material the nonlinear optical response of the medium also behaves on the average as a homogeneous with an effective dielectric tensor ϵ. To calculate the components of this tensor, a statistical theory has been developed. For each medium we assume a Kerr nonlinearity

$$\epsilon = \epsilon_{mL} + \chi_m^{(3)} E_m(\omega) E_m(\omega) \quad , \tag{4}$$

for the embedded metal particles and the host medium has a linear dielectric function ϵ_h. For an isotropic effective medium, the dielectric function can be expanded as a power series in the applied electromagnetic field, E_O:

$$\overline{\epsilon} = \overline{\epsilon}_L + \overline{\chi}^{(3)}\left|\vec{E}_o\right|^2 ; \tag{5}$$

where $\overline{\chi}^{(3)}$ is the effective medium Kerr coefficient.[3,5] For small volume fraction of metal particles, $f \ll 1$, its value is:

$$\overline{\chi}^{(3)} \approx f \, \chi_m^{(3)} \left|\gamma^m\right|^2 \left(\gamma^m\right)^2 . \tag{6}$$

The expression has the same enhancement factor with the surface plasmon resonance that appears in the absorption coefficient in Eq.(3). The surface-plasmon resonance is defined by the vanishing of the real part of the denominator in Eq.(2). This enhancement is proportional to the fourth power of the resonance.

A DFWM experiment is a direct method for observing the effects of the Kerr nonlinearity $\overline{\chi}^{(3)}$. The counter-propagating pump fields are denoted as $E_F(o)$ and $E_B(L)$; the length of the interaction region is L. The conjugate reflectivity with absorption effects is seen as:

$$R_c = \frac{\left|E_c(0)\right|^2}{\left|E_s(0)\right|^2} = 4\left|\beta\right|^2 \frac{e^{-2\alpha L}\left|\sinh\left(\frac{\alpha L}{2}\right)\right|^2}{\alpha^2 L^2}\left|E_F(0)\right|^2\left|E_B(L)\right|^2 , \tag{7}$$

101

where the coefficient $\propto$ is the linear absorption coefficient. The value of β is

$$\beta = \bar{\chi}^{(3)} \frac{3}{2} \frac{1}{\sqrt{\epsilon_h} \left(\epsilon_m''\right)^2} \left(\frac{c}{\omega}\right). \tag{8}$$

where c is the speed of light. Note in Eqn. 8 that β is frequency dependent.

To summarize the findings of previous researchers [3,5], the presence of the surface plasmon resonance can be more sensitively probed in the conjugate reflectivity because it is proportional to the square of the effective Kerr nonlinearity $\chi^{(3)}$ and hence, small changes in the enhancement are magnified by an eighth power of the resonant denominators. The absorption modifies the conjugate reflectivity to such a degree that the optimum value of the reflectivity at any given frequency occurs for $\propto L$ approximately unity; for larger values, obtained either by increasing the interaction length L or increasing the concentration of particles, the reflectivity is decreased due to the depletion of the field amplitudes.

In the present experimental arrangement, Eq.(7) must be modified to include the absorption of the waves before they have reached the interaction volume (see Fig. 2). Each of the wave amplitudes will experience a loss that is summarized by multiplying Eq.(7) by a factor:

$$exp\left(-\propto L'\right).$$

The length L' is a sum over all the lengths from each of the four wave amplitudes outside the interaction region. Eq.(9) has the effect of reducing the reflectivity at the resonance frequencies where the absorption is highest. The observed conjugate reflectivity at the surface plasmon resonance can be greatly suppressed by the absorption of the wave amplitudes outside the interaction volume. Specific results of the theory will be discussed in the following sections.

SAMPLE FABRICATION

The gold microparticles were prepared following a recipe by Perenboom et al [6]. The method consists of adding a reducing agent to a solution of metal ions causing condensation of the gold. The colloidal solution had a gold volume fraction of 5×10^{-6}. The recipe predicted the gold particles to have a mean diameter of 15nm with a standard deviation of <2 nm. We examined the particles in a TEM capable of 0.2 nm resolution and found good agreement with Perenboom's results on particle sizes. The solution remained stable for months at room temperature without any noticeable aggregation of the individual particles. Conglomerations of particles cause a red shift in the surface plasmon resonance and a decrease in the absorbance peak [7].

Figure 1 shows the linear absorbance ($A = -log_{10}$Transmittance) of the colloidal gold in a spectrophotometer cell with a 5 mm path length. The resonance observed at 525 nm is due to

surface plasmon excitation in the gold particles [6]. Also plotted in Fig. 1 is the absorbance of undoped gelatin (purified gelatin from Fisher Scientific) in a 5 mm cell. The gelatin sample was made by mixing 18 ml of water and 2.5 gm of gelatin on a hot plate.

Gelatin waveguides were made on 2.5x5 cm^2 pieces of Corning 7059 glass by a spin coating process. A mixture of 20 ml of water and 2.2 gm of gelatin was stirred over low heat until the gelatin granules dissolved. After 5 minutes of cooling, the solution was flooded on a glass slide and spun at 700 rpm for 20 seconds. The sample was immediately placed in an oven at 90°C for 3 minutes to remove the excess water. The gelatin waveguides used in the experiment were 1.8±0.1μm thick. The resulting films were quite stable but did have some defects presumably due to small voids in the film caused by trapped air. We measured the index of the gelatin to be 1.542 and the index of the substrate to be 1.528 at $\lambda = 633$ nm. The waveguides were single mode at 633 nm. At shorter wavelengths near 530 mm, a weak second mode was observed.

The gelatin waveguides were doped with gold particles by substituting the colloidal solution for water in the gelatin mixture. Three different doping levels were examined in the DFWM experiment. The mixtures of colloidal solution (ml), water (ml), and gelatin (gm), were Sample I: 0.55, 19.45, and 2.2; Sample II: 1.1, 18.9, and 2.2; Sample III: 2.2, 17.8, and 2.2 respectively. The volume fractions of gold in the waveguide were estimated to be $1.25x10^{-6}$, $2.5x10^{-6}$, and $5x10^{-6}$. The estimates were based on the known amounts of gold and gelatin in each sample and on propagation losses in the waveguide. The true values of the doping fractions are expected to be within 30% of the estimated values. Typical propagation losses for an undoped gelatin waveguide were slightly less than 1dB/cm at 585 nm.

A waveguide with a volume fraction of gold $f = 5X10^{-6}$ had propagation losses of ~ 6dB/cm at 585 nm.

In order to make a direct comparison of metal particle composites and semiconductor composites, an ion exchanged waveguide was fabricated in a Corning 3-69 glass following the process outlined in Ref. 2. The Corning filter glasses have a $\overline{\chi}^{(3)} = 10^{-8}$ esu [8] and a volume ratio of CdS_xSe_{1-x} to glass of $f=10^{-3}$ [19].

EXPERIMENT

An excimer pumped dye laser provided 10 ns pulses at a 10Hz repetition rate. The tuning range for the dye was 530-590 nm. This wavelength range extends from the peak in the absorbance spectrum of Au particles out to the wing where the absorbance is relatively low. A standard DFWM setup was used for the measurements [8], two counterpropagating pumps of equal intensity and a weak probe incident at an angle of five degrees with respect to the forward pump. The conjugate signal was detected by a Si APD. The pulse energy was measured by a pyroelectric detector with an amplifier stage. The frequency dependence of the optical components and detectors was measured in order to correct the raw data.

The three beams were prism coupled into the TM mode of the waveguide (Fig. 2). The probe and forward pump were coupled into the waveguide through one prism and the backward pump was coupled through a second prism. The prism coupling efficiencies were ~7%. We did not observe a significant change in the prism coupling efficiency over the range of pump intensities used in the experiment. This can be attributed to the relatively small change in refractive index with pump intensity in addition to variations of the gelatin film thickness which leads to a range (albeit small) of coupling angles.

Ideally, we would like to have maximum overlap of the beams in the waveguide to minimize propagation losses before the pulses reach the interaction volume. This can be achieved by having a large beam diameter and short interaction lengths along with a small angle between the probe and forward pump. The ideal situation is not always possible due to mounting constraints and limited laser power. These propagation losses will tend to skew reflectivity versus frequency plots if the absorbance is strongly frequency dependent. Experimental results of the reflectivity as a function of wavelength presented below were made at a constant pump intensity of $8MW/cm^2$. The prism coupling angle was adjusted to maximize the conjugate signal at each wavelength.

RESULTS

Figure 3 displays a typical nonlinear response of a gelatin waveguide doped with gold particles. On a log-log plot, the conjugate reflectivity versus pump intensity has a slope of two which is characteristic of a third order nonlinearity. Also plotted in Fig. 3, the reflectivity of a waveguide made in a SDG (Corning 3-69). The conjugate reflectivities of the gold doped waveguides and SDG waveguides are comparable to the reflectivities from the bulk material. For a given pump intensity the reflectivity of the SDG is higher than the reflectivity of the gold doped waveguide. The difference in the reflectivity is primarily attributed to the higher volume of semiconductor to glass ($f = 10^{-3}$) compared with the volume, fraction of gold to gelatin ($f = 10^{-6}$). The reflectivity of the SDG saturates at $1MW/cm^2$ while no saturation was noted for the gold doped waveguide. However, at pump intensities $>50MW/cm^2$ damage was observed in the gelatin waveguides. From Fig. 3 we estimate that a gold particle has a value of $\chi_m^{(3)} \approx 1 \times 10$-8 esu which is consistent with the value reported in Ref. 3.

An undoped gelatin waveguide was placed in the DFWM setup to insure the conjugate signal resulted from the gold particles. Although a signal was not detected from pure gelatin waveguides, this null result does not necessarily imply that the gold particles have an intrinsic nonlinearity. Associated with the surface plasmon excitation are large field enhancements both inside and outside the gold particles. Small nonlinearities in the host can be magnified by the enhancement of the applied field. Local field enhancements outside metal particles are responsible for the large Raman signals in surface enhanced Raman scattering [11]. A calculation of the "enhanced nonlinearity" of the host is given in Ref. 12. The enhancement can be quite large but it occurs over a limited volume about the sphere. When weighted by the

low volume fraction of 10^{-6}, the overall enhancement of the nonlinearity in the host is negligible.

The conjugate signal as a function of wavelength is plotted in Fig. 4 for various doping levels. While the shape of the curves in Fig. 4 are representative of each doping level, there was some variation in the magnitude of the signal from samples with equal volume fractions. The signal variations are presumably due to differences in film quality resulting from the spin coating process. Defects were avoided whenever possible by translating the sample with respect to the incident beams.

At a volume fraction of $f = 1.25 \times 10^{-6}$ the conjugate signal follows the local field enhancement factor and peaks in the vicinity of the surface plasmon resonance. As the volume fraction is increased to $f = 2.5 \times 10^{-6}$, absorption begins to have a detrimental effect on the conjugate signal. The optimum operating wavelength moves away from the resonance into a region of less absorption. If the volume fraction is doubled once more to $f = 5 \times 10^{-6}$, the signal at 530 nm continues to degrade and the optimum wavelength moves out beyond 580 nm. The reflectivity plotted in Fig. 3 for a gold doped waveguide was measured at $\lambda = 560$ nm and $f = 5 \times 10^{-6}$. By proper choice of operating wavelength and doping fractions, the reflectivity plotted in Fig. 3 could be increased by at least a factor of three. Better overlap of the beams in the sample would also improve the signal, especially at the plasmon resonance where absorption is at a maximum.

Figure 5 is the theoretical response of the waveguides for various doping levels of gold assuming a frequency independent $\chi_m^{(3)}$. In analysing the experimental data on the conjugate reflectivity, we use the experimental absorbance curve adjusted for each concentration as shown for one case in Fig.(1). The factors of $|\gamma^m|^8$ in Eq.(7) were replaced by the fourth power of the experimental absorption using Eq.(3). The frequency dependence of the other factors are important and must also be retained. The frequency dependence of ϵ''_m was taken from the data of Palik.[13] The idealized case of complete beam overlap inside the sample, $L' = 0$, is plotted in Fig. 5a. As the volume fraction of gold increases, the peak in the reflectivity experiences a red-shift due to increasing absorption losses. Fig. 5b is the calculated reflectivity for a more realistic case in which the laser beams in the DFWM experiment undergo additional losses due to incomplete overlap of the beams in the sample, $L' = 5$ mm. The overall reflectivity decreases by $\sim 2x$ compared to the case in which $L' = 0$. Also, a stronger red-shift of the maximum reflectivity is observed. The results of Fig. 5b are in qualitative agreement with the experimental data of Fig. 4. Discrepancies between theory and experiment in the relative magnitudes of the peaks are likely due to variations in film quality of the samples.

The response time of the gold colloids was investigated by inserting a delay line in the apparatus. Fig. 6 is a plot of the conjugate signal versus delay of the backward pump beam. The conjugate signal is observed to decay with time, however, a substantial signal is

evident for delays up to three pulse durations (30 ns). Similar behavior was noted for the Corning SDG although the signal decayed slightly faster compared with gold particles

Originally, $\chi_m^{(3)}$ of metal particles was attributed to quantum size effect [3]. The value of $\chi_m^{(3)}$ was predicted to be proportional to $1/r^3$ when r is the radius of the gold particle. Recent measurements on colloidal gold solutions with sizes ranging from 5-30 nm found no evidence to support a size dependent $\chi_m^{(3)}$ [12]. Similar measurements on gold particles in a glass matrix also found $\chi_m^{(3)}$ to be size independent [15]. A possible mechanism for the nonlinearity is Fermi smearing [14,15]. The incident pulse changes the population of states about the Fermi energy which affects the transition probability of d-band electrons to states near the Fermi energy. The rise time is subpicosecond, with a decay time limited by the cooling time of the electron gas. The electron gas cools to the gold lattice through electron-phonon scattering in 2-3 ps. For the Fermi smearing mechanism, the local field enhancement factor of Eqn. 2 remains an integral part of the overall $\chi^{(3)}$ as defined in Eqn 6. The experimental results of Refs. 12 and 15 are consistent with a Fermi smearing mechanism.

Ref. 12 presents measurements of the response time of gold particles as a function of particle size. At a 10 Hz rep. rate, 30 ps pulses, and 50MW/cm^2 pump intensity the 5 and 10 nm diameter particles have a response time limited by the pulse duration. For 15 nm diameter particles a small slow component appears on the response curve and for 30nm diameter particles the slow component is obvious. The size dependent slow component is attributed to a thermooptic nonlinearity due to lattice heating of the gold particle. The cooling time of the gold lattice by diffusion to the host depends on the surface area of the particle. For a 15 nm gold particle in water the lattice cooling time is 30 ps. The relative magnitude of the nonlinearity due to Fermi smearing and lattice heating will depend on the particle size, the pulse duration, the pulse energy, and the pulse rep. rate.

The time response plotted in Fig. 6 indicates the possibility of a mechanism at least as fast as the 10 ns pulse duration and a slower response which is too slow to originate from lattice heating of the gold particle. A likely mechanism is a thermooptic nonlinearity due to heating of the host. The decay time will then be limited by the thermal conductivity of the host.

In summary, planar waveguides have been fabricated with a nonlinear guiding medium of gelatin doped with gold particles. DFWM measurements show the conjugate reflectivity in waveguide geometry was comparable to that of the bulk material. The response of the samples as a function of frequency and doping levels was examined theoretically by an effective medium theory. The results have shown that absorption is important for doping levels of only 10^{-6}. It should be possible to increase the conjugate reflectivity of metal particle composites by utilizing the local field enhancement. Local field factors can be increased by using nonspherical particles which exhibit the "lightning rod" effect (large fields near surfaces of high

curvature) [16]. Another possibility would be to encase a nonlinear microparticle in a thin shell of metal or vice versa which leads to enhancement factors of 10^8 in certain cases [17]. As pointed out in the present paper, absorption will be the limiting factor in achieving higher conjugate signals.

REFERENCES

[1] G. I. Stegeman, E. M. Wright, N. Finlayson, R. Zanoni, and C. T. Seaton, J. Lightwave Tech. 6, pp. 953-970 (1988).

[2] A. Gabel, K. W. DeLong, C. T. Seaton, and G. I. Stegeman, Appl. Phys. Lett. 51, pp. 1682-84 (1987).

[3. F. Hache, D. Richard, and C. Flytzanis, J. Opt. Soc. Am B 3, pp. 1647-55 (1986).

[4] M. J. Bloemer, M. C. Buncick, R. J. Warmack, and T. L. Ferrell, J. Opt. Soc. Am B 5, 2552-2559 (1988).

[5] J. W. Haus, N. Kalyaniwalla, R. Inguva, M. Bloemer, and C. M. Bowden, J. Opt. Soc. Am B 6, 797-807 (1989); J. W. Haus, R. Inguva, and C. M. Bowden, Phys. Rev. A, in review (1989).

[6] J. A. A. J. Pereboom, P. Wyder, and F. Meier, Physics Reports 78, pp. 173-292 (1981).

[7] M. Quinten, U. Kreibig, and D. Schonauer, Surface Science 156, pp. 741-750 (1985).

[8] R. K. Jain and R. C. Lind, J. Opt. Soc. Am 73, pp 647-653 (1983).

[9] D. W. Hall and N. F. Borrelli, J. Opt. Soc. Am. B 5, pp. 1650-1654 (1988).

[10] M. Born and E. Wolf, "Principles of Optics," New York: Pergamon Press, Fifth Ed. (1975); see section 13.5 on the theory of Mie.

[11] A. Wokaun, Solid State Physics 38, pp. 223-294 (1984).

[12] M. J. Bloemer, J. W. Haus, and P. R. Ashley, accepted for publication in J. Opt. Soc. Am. B (1990).

[13] E. D. Palik, ed. Handbook of Optical Constants of Solids (Academic, New York, 1985).

[14] R. W. Schoenlein, W. Z. Lin, J. G. Fujimoto, and G. L. Eesley, Phys. Rev. Lett. 58, pp. 1680-83 (1987).

[15] F. Hache, D. Ricard, C. Flytzanis, and U. Kreibig, App. Phys A 47, pp. 347-357 (1988).

[16] P. F. Liao and A. Wokaun, J. Chem. Phys. 76, pp. 751-752 (1982).

[17] A. E. Neeves and M. H. Birnboim, Optics Lett. 13, pp. 1087-1089 (1988).
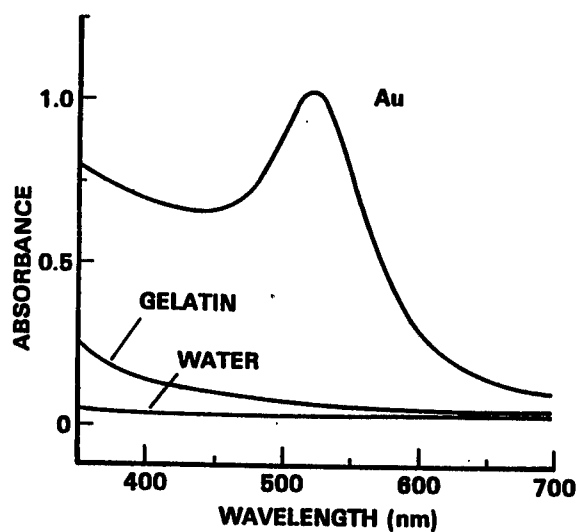
Figure 1. Linear absorbance spectra of colloidal gold, undoped gelatin, and water in a spectrophotometer cell with a 5 mm path length. The spectra include the absorbance of the cell.
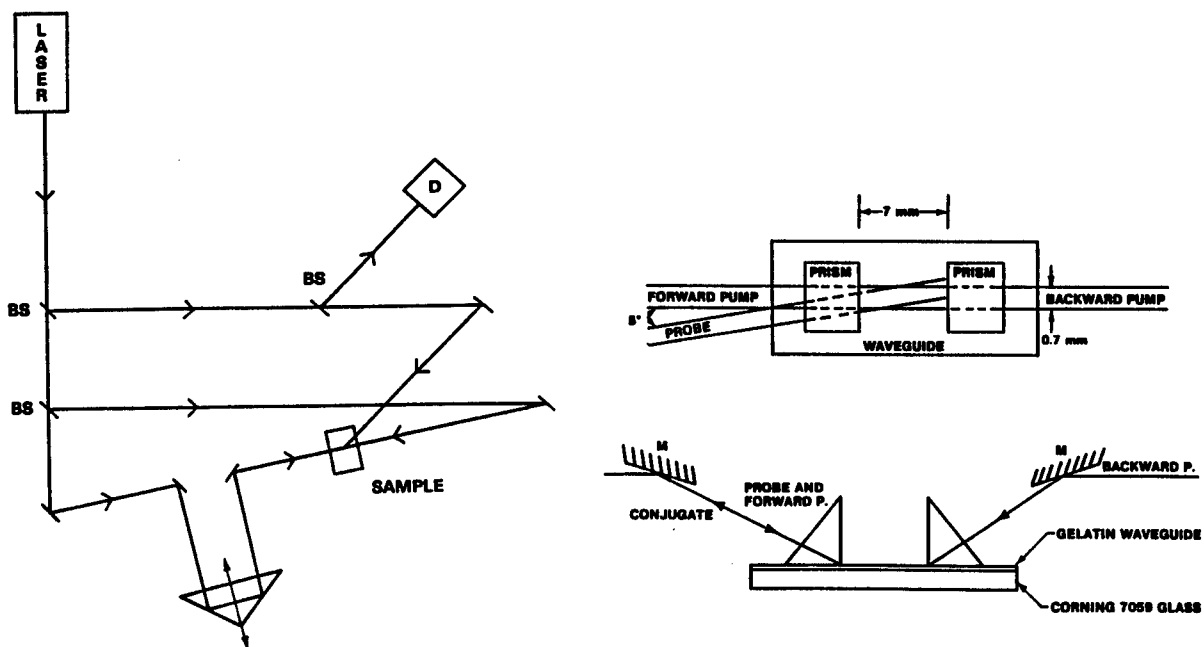
Figure 2. Diagram of experimental arrangement. Not to scale.
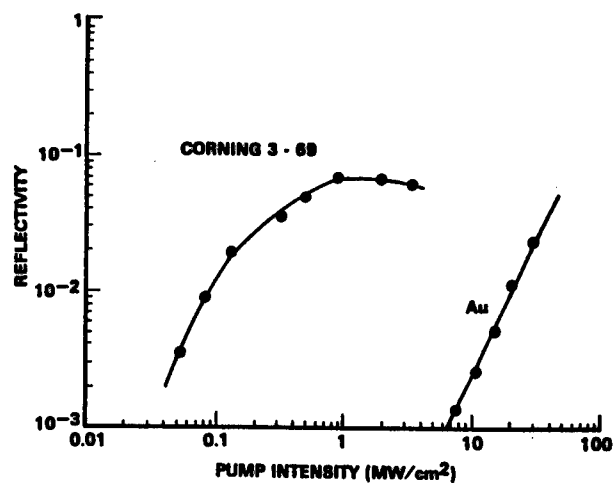
Figure 3. Conjugate reflectivity versus pump intensity on a log-log plot of a gelatin waveguide doped with gold particles and of an ion-exchanged waveguide made in a semiconductor doped glass. The measurements were made at λ= 532 nm for the Corning 3-69 and λ= 560 nm for the gold doped waveguide. The volume fraction of gold to gelatin was f = 5×10^{-6}.
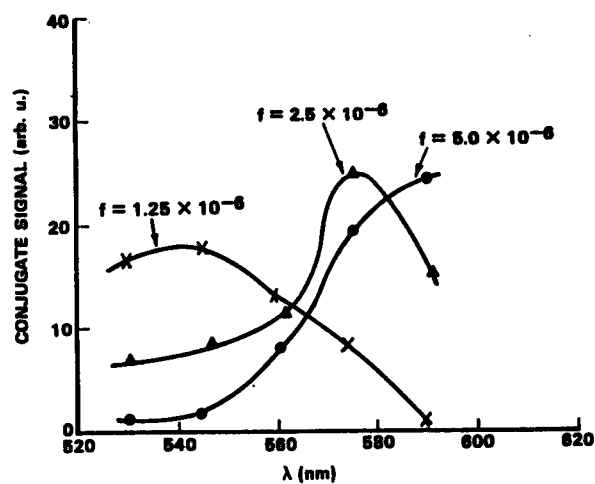
Figure 4. Conjugate signal as a function of wavelength at a constant pump intensity for gelatin waveguides doped with varying amounts of gold particles.
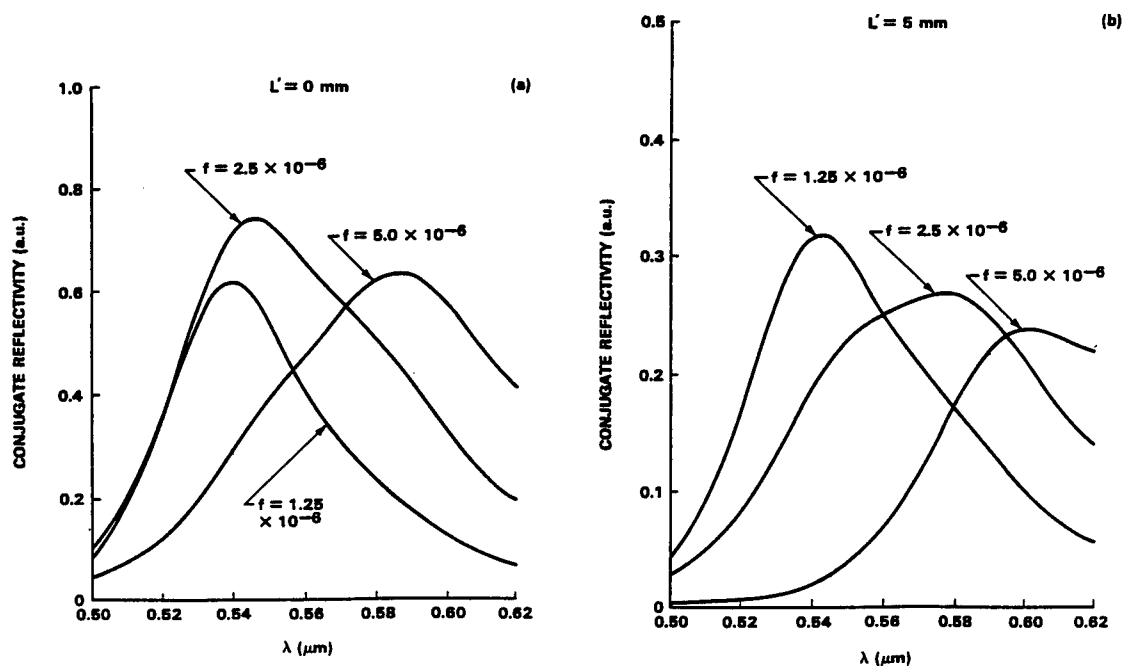
Figure 5. Theoretical conjugate reflectivity versus wavelength for gold filling fractions of $f = 1.25 \times 10^{-6}$, 2.5×10^{-6}, and 5×10^{-6}; (a) assuming complete overlap of the beams inside the waveguide, $L' = 0$, (b) assuming each of the four beams experience some loss in the waveguide which occurs outside the interaction region, $L' = 5$mm.
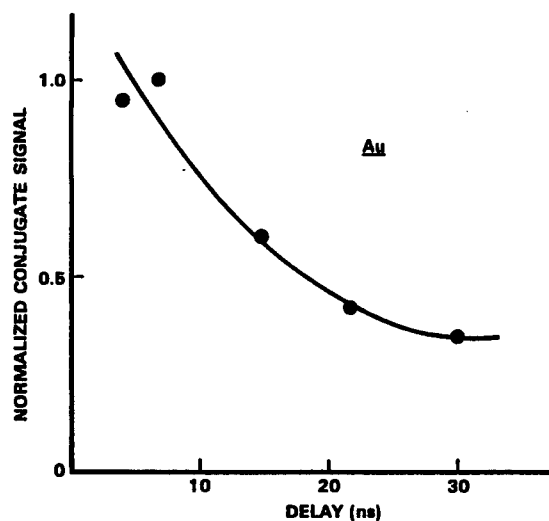
Figure 6. Conjugatge signal from gold particles versus time delay of the backward pump at $\lambda = 532$ nm and a pump intensity of 8 MW/cm^2. The pulse duration was 10 ns.

Nonlinear Optical Properties of Metallic Microparticle Composites (U)

Charles M. Bowden, Dr, and *Mark J. Bloemer, Dr.
Research Directorate, AMSMI-RD-RE-OP
U.S. Army Missile Command, Redstone Arsenal, AL 35898-5248

and

Joseph W. Haus, Dr.
Physics Department, Rensselaer Polytechnic Institute
Troy, NY 12180-3590

1. INTRODUCTION

Light scattering in inhomogeneous media has been an active research area with a broad range of problems[1]. Recent work on nonlinear effects has established conditions of resonant enhancement of these nonlinearities and sought to use them to increase the magnitude of physical effects, such as optical phase conjugation,[2,3] or to introduce new phenomena, such as intrinsic optical bistability.[4,5] Of course, this research has its roots in the studies of surface-enhanced Raman scattering[6], where, for example, the Raman cross section observed from molecular monolayers upon a silver substrate was enhanced by several orders of magnitude.

The enhancement of the Raman scattering cross section is due to the large increase of the local field, both inside and outside the metal particles near its surface, at the surface-plasmon resonance. Silver has a sharp resonance feature, but the resonance in gold and other metals can also lead to strong Raman scattering.

We predict that intrinsic optical bistability should be observed in silver particle composites. This optically bistable phenomenon is a local condition in the material and occurs without optical feedback, as from a cavity. The bistable behavior is due to the local field effect associated with the particles embedded in the dielectric host material. For spherical particles, the switching intensities are expected to be ~ 200 MW/cm^2; this magnitude is determined by the magnitude of the imaginary part of the metal's dielectric constant[5,7]. The

113

design of experiments with ellipsoidal particles and a nonlinear host medium for the particles can lead to a significant reduction of the switching intensities. This is possible because the resonance frequency can now be chosen so as to minimize the imaginary part of the dielectric constant.

Degenerate four-wave mixing (DFWM) measurements in metal colloids reveal that a small silver or gold sphere has $\chi^{(3)} = 10^{-8} - 10^{-9}$ esu in the vicinity of the surface plasmon resonant with response times of the order of picoseconds.[2,3] It has been proposed that the $\chi^{(3)}$ for small metal particles results from an intrinsic $\chi_m^{(3)}$ of the plasmon excitation. The conjugate signal is proportional to the local field factor to the eighth power. The local field factor's dependence on shape has been illustrated many times in surface-enhanced Raman scattering experiments. In addition, it is possible to shift the resonant energy of the surface plasmon by varying the shape of the particle.[8] We found that it was important to make a more detailed examination of these enhancement effects. We present results from studies of the linear absorption as well as of the conjugate signal in DFWM experiments. Analysis of these experiments gives detailed information about the particle shapes.

Using results from the effective-medium theory[7], we evaluate the effects of absorption, both linear and nonlinear, and the particle shape on the DFWM conjugate signal.

2. RESULTS FROM THE EFFECTIVE-MEDIUM THEORY

In a composite dielectric material that has metal particles dispersed in it, the medium behaves on the average as a homogeneous medium with an effective dielectric tensor $\bar{\epsilon}$. To calculate the elements of this tensor, statistical theories have been developed that examine the effects of multiple scattering of the applied electromagnetic fields on linear media.[7,9-12] More recently, these theories have been extended to cover nonlinear systems.[7] We follow the results of Ref. 7, from which the effective dielectric tensor can be determined through the self-consistency condition

$$0 = \left\langle \left\{ I + \left[\epsilon(x) - \bar{\epsilon} \right] \cdot \Gamma \right\}^{-1} \left[\epsilon(\vec{x}) - \bar{\epsilon} \right] \right\rangle , \tag{1}$$

where I is the unit tensor, $\epsilon(\vec{x})$ is the inhomogeneous dielectric tensor of the medium, and Γ is the depolarization tensor. The average in Eq. (1) contains uniform weighting of the particle positions, as is consistent with their random positions, and also an average over the weighted orientations of the particles.

The dielectric constant in the inhomogeneous medium is that of either the host dielectric matrix, $\epsilon_h(\vec{E_L})$, or the metallic particles, $\epsilon_m(\vec{E_L})$. Both may be nonlinear functions of the electric field, but the field is not identical for each environment. Outside the conducting particles, the field is the applied field plus the induced dipole field. Inside the particles, the local field drives the optical nonlinearities of that medium. For each material we assume a Kerr medium, which for degenerate field is written as

$$\epsilon_h = \epsilon_{hL} + \chi_h^{(3)} : \vec{E}_h(\omega) \vec{E}_h(\omega) \tag{2a}$$

for the host medium and

$$\epsilon_m = \epsilon_{mL} + \chi_m^{(3)} : \vec{E}_m(\omega) \vec{E}_m(\omega) \tag{2b}$$

for the embedded material. In general, the coefficients $\chi_\propto^{(3)}$ are fourth-rank tensors.

The linear dielectric tensors are denoted by a subscript L. We will make the further assumption below that these tensors have degenerate principal values, i.e., they are isotropic media. This does not imply that the effective-medium dielectric functions are isotropic; for ellipsoidal particles the angular distribution of the scattered light depends on the particles' shape and orientation. When the particle orientations are correlated, this is manifest by polarization dependence in the scattered light. For gold particles of dimensions < 30 nm the linear optical properties are dominated by absorption, with < 10% of the total cross section being due to scattering. Spheroidal gold particles support two dipole modes with a moment aligned along the minor axis of the particle. Therefore the frequency dependence of the absorption cross section for polarized light is sensitive to the spheroids' orientation. The effective-medium theory can incorporate this polarization difference by an anisotropic dielectric tensor for the effective medium.

The field inside the ellipsoidal-shaped particles is uniform but not necessarily in the same direction nor of the same magnitude as the applied field $\vec{E}_0$. For the ellipsoids of metal particles embedded in an effective medium, we have the local field

$$\vec{E}_m(\omega) = \gamma^m(\omega) \cdot \vec{E}_0(\omega) \; , \tag{3a}$$

and for ellipsoids of the host material we have its local field

$$\vec{E}_h(\omega) = \gamma^h(\omega) \cdot \vec{E}_0(\omega) \; . \tag{3b}$$

The local fields may be quite different from the applied field. Their value is determined by the tensors γ^α, where $\alpha = m$ or $\alpha = h$; these are related to the depolarization tensor and the dielectric tensors:

$$\gamma^m(\omega) = \left\{ I + \Gamma(\omega) \left[\epsilon_m(\omega) - \bar{\epsilon}(\omega) \right] \right\}^{-1} \tag{4a}$$

and

$$\gamma^h(\omega) = \left\{ I + \Gamma(\omega) \left[\epsilon_h(\omega) - \bar{\epsilon}(\omega) \right] \right\}^{-1} \; . \tag{4b}$$

Since they provide the magnitude of the enhancement effects of the local field, we will denote these coefficients as *enhancement factors*. These factors determine the resonance position in the denominator, which is controlled by the principal values of the depolarization tensor.

For random orientation of the spheroids, the enhancement factors in Eqs. (4), which refer to a coordinate system fixed to the spheroid, now have different contributions because the applied field is changing its direction relative to this coordinate system. The lack of a single local field for particles with different orientations means that the self-consistency condition, Eq. (1), incorporates a complicated angular dependence in the nonlinear coefficients. A simple expression for the effective-medium dielectric function is no longer possible; instead the effective medium is expanded as a power series in the applied field.

A useful expression for the nonlinear coefficient $\bar{\chi}^{(3)}$ is obtained when the host medium is linear, $\chi_h^{(3)} = 0$, and for small concentrations the result is

$$\overline{\chi}^{(3)} \simeq f \frac{\chi_m^{(3)}}{\epsilon_{hL}} \left\{ \frac{\frac{8}{15}\left|\gamma_x^m\right|^2 + \frac{2}{15}\left|\gamma_z^m\right|^2}{\left[\left(1-A_x\right)\overline{\epsilon}_L + A_x\epsilon_{mL}\right]^2} + \frac{\left(\frac{2}{15}\left|\gamma_x^m\right|^2 + \frac{3}{15}\left|\gamma_z^m\right|^2\right)}{\left[\left(1-A_z\right)\overline{\epsilon}_L + A_z\epsilon_{mL}\right]^2} \right\} . \quad (5)$$

Here, A_x and A_z are depolarization factors, i.e., $A_x = \overline{\epsilon}_x \Gamma_{xx}$, $A_z = \overline{\epsilon}_z \Gamma_{zz}$.

This expression has denominators with two surface-plasmon resonances. The strongest resonances occur when the enhancement factors in the numerator have their resonance at the same frequency as the denominator, as they are in the first and last terms of Eq. (5). However, because of the mixing of these resonant terms, enhancements of the susceptibility occur between the frequencies. At the surface-plasmon resonances the effective nonlinearity can be enhanced by several orders of magnitude over the value $\chi_m^{(3)}$. The actual enhancement factor depends on the metal. In silver the resonance is sharp in the visible, and 8 to 10 orders of magnitude enhancement are possible. This is reduced by the volume fraction, f, which multiplies these factors.

The conjugate reflectivity is defined as

$$R = \frac{\left|E_c(0)\right|^2}{\left|E_s(0)\right|^2} = 4\left|\beta\right|^2 \frac{e^{-2\alpha L}\left|\sinh\left(\frac{\alpha L}{2}\right)\right|^2}{\alpha^2} \left|E_F(0)\right|^2 \left|E_B(L)\right|^2, \quad (6)$$

where $\beta = 3\omega^2 \overline{\chi}^{(3)}/(2c^2 k)$ and $\alpha = \omega^2 \operatorname{Im}(\overline{\epsilon})/(c^2 k)$ are the nonlinear coupling and absorption coefficients, respectively. Here E_F and E_B are the forward and backward pump field amplitudes, respectively, and $E_c(0)$ and $E_s(0)$ are the conjugate and probe fields of the input.

A more accurate expression valid for large nonlinear coupling can be found in Ref. 13, but these expressions are not needed for the discussion that follows.

Equation (6) contains $|\overline{\chi}^{(3)}|^2$, which contains the enhancement factors to the eighth power. The absorption coefficient, α, contains the first power of the enhancement factor; therefore the overall reflectivity has a sixth power of the enhancement for the multiplicative coefficient of the hyperbolic sine.

3. THEORETICAL RESULTS

The description of heterogeneous nonlinear-optical materials given in the previous sections is useful for a wide range of materials; these include semiconductor colloids and glasses and, under certain restrictions on the size, heterogeneous polymer solutions. In nonconducting media, the transparency is usually large enough that high-volume fractions can be considered.

The pump fields are taken as equal and small, such that $\chi_m^{(3)} |E_B(0)|^2 = 0.01$.

In Fig. 1 the conjugate reflectivity for spheroidal particles is shown from Eq. (6). The length of the medium is taken as L = 2 nm; the maximum conjugate reflectivity occurs at a value $\alpha L \simeq 1.5$. At small concentrations only a single maximum is observed, but the high concentrations show the appearance of a second peak, which eventually dominates the reflectivity. Over a range of concentrations the reflectivity changes from a single peak to a broad double-peaked structure. A second peak at the second surface plasmon resonance eventually dominates the reflectivity.

The second peak occurs because $\overline{\chi}^{(3)}$ is enhanced at this frequency owing to the second surface-plasmon resonance, and the absorption has not yet become the limiting factor in the maximum of the conjugate reflectivity. The peak in R_c can switch from one resonance to another; this provides a sensitive test of the average particle shape.

4. EXPERIMENTAL RESULTS

We examined the linear- and nonlinear-optical properties of glass containing gold particles. The glasses are RG6 filters obtained from Schott. The volume fraction of gold in the RG6 is fixed; therefore we varied the sample thickness to investigate the effects of absorption on the DFWM signal. Three samples of 1-, 2-, and 3-mm thicknesses provided αL = 1.6, 3.1, 4.7, respectively, at λ = 0.545 μm. Recall that the maximum conjugate reflectivity occurs at $\alpha L \simeq 1.5$. Nearly all the loss in the RG6 glass is due to the gold particles, and little is due to the glass matrix.

118

DFWM was performed at different frequencies using an excimer-pumped dye laser at a 10-Hz repetition rate. The pulse duration was 12 nsec for all frequencies. We do not have electron micrographs of the gold particles in the RG6 glass, but it is not unreasonable to assume that the particles' shape may deviate from a perfect sphere. The particles would also be expected to have a range of shapes in the actual glass.

Figure 2 shows the conjugate signal versus wavelength at a constant pump intensity of 8 MW/cm^2. For the thinnest sample, L = 1 mm, one broad peak is observed. The thickness L = 2 mm shows an overall reduction of the reflectivity and two peaks in the spectrum, which is characteristic of the two surface-plasmon resonance frequencies supported by a nonspherical particle. For L = 3 mm, the dominant peak in the reflectivity has shifted to the longer wavelength, and the overall reflectivity has been reduced by a factor of ~ 3 compared with the 1-mm sample. A dip in the conjugate signal versus wavelength is predicted for spherical particles but only at values of αL larger than those of the RG6 samples examined here.

5. OPTICAL BISTABILITY

Optical bistability without an optical cavity [called instrinsic optical bistability (IOB)] has been observed in a number of materials[14] and theoretical descriptions have been given for this phenomenon[15]. Recently, this IOB phenomenon has been predicted for a single-semiconductor microparticle[16]. We extend these predictions by considering the effects of a composite consisting of a collection of conducting microparticles randomly distributed throughout a transparent host material. The particles are considered to possess a sharp size distribution.

To be specific, we consider a material made up of small spherical metallic particles of radius a, with volume fraction f, and they are embedded in a dielectric host medium with coefficient ϵ_d. The metal grains are small enough that surface effects are important and are driven by a strong applied field so that they will have a nonlinear response which turns out to be of the Kerr type for gold and silver particles.

A material composed of microscopic particles that are randomly and independently distributed throughout the medium has a dielectric constant that varies from point to point in the medium $\epsilon(r)$. The medium is described by an effective dielectric function ϵ^*, which accounts for the multiple-scattering effects of the heterogeneous medium in an averaged manner. The expression relating ϵ^* to the microscopic properties of the medium is determined by the self-

119

consistency condition. In our case, this general expression involves the volume fraction f and the dielectric constants. However, the concentrations of metal spheres we will explore are low enough that the effective dielectric function for the medium can be approximated by linear terms in the concentration (Maxwell-Garnett approximation)

$$\epsilon^* \approx \epsilon_d + f \left[3\epsilon_d / \left(2\epsilon_d + \epsilon_m \right) \right] \left(\epsilon_m - \epsilon_d \right) . \tag{7}$$

The steady-state Maxwell equation for the applied, propagating electric field is $[E(r,t) = E_0(r) e^{-i\omega t}]$

$$\nabla^2 E_0 + \left(\omega^2/c^2 \right) \epsilon^* E_0 = 0. \tag{8}$$

To develop the theory further, first we introduce the slowly varying envelope approximation and use one-way propagation of the electromagnetic field. To be specific we assume the electromagnetic field is propagating along the positive z axis:

$$E_0 = E_\epsilon e^{ikz} . \tag{9}$$

The wave number k is chosen to eliminate the real, linear contribution to the dielectric function ϵ^* in Eq. (7),

$$k^2 = \left(\omega^2/c^2 \right) \text{Re} \left\{ \epsilon_\ell^* \right\} , \tag{10}$$

where $\epsilon_\ell^* = \epsilon^*$ ($E_L = 0$). The transverse effects are neglected; their contribution can be important for small Fresnel numbers as we have found for the nonlinear oscillator model, but the results for the on-axis intensity were not affected in that study down to Fresnel numbers of unity. As the two problems are quite analogous, we expect the same results for this model. The backward-propagating wave is negligible in the present case because we restrict our numerical results to small concentrations of metal spheres, $f \ll 1$. In this limit the nonlinear changes of the real part of dielectric function are small. The Maxwell field equation (8), in a slowly varying envelope approximation (SVEA), is reduced to

$$\frac{dE_\epsilon}{dz} - i\kappa E_\epsilon = 0 .$$ (11)

The coefficient appearing in Eq. (11) is

$$\kappa = \left[\epsilon^* - \mathrm{Re}\left(\epsilon_\ell^*\right)\right] \omega/2c \sqrt{\mathrm{Re}\left(\epsilon_\ell^*\right)}$$ (12)

These equations provide the basis for studying propagation effects in composite media.

In the following we take the spherical particles to be composed of silver and use the dielectric constant which was developed for small particles to include the quantum-mechanical confinement of the electrons. The linear dielectric constant in Eq. (5) is given by

$$\epsilon_0 = \epsilon_\infty + \left[\omega_p^2 / \left(\Omega^2 - \omega^2 - i\omega\Gamma\right)\right] .$$ (13)

The coefficient ϵ_∞ = 4.66. The resonance frequency Ω^2 and the damping coefficient, Γ , are functions of the particle size:

$$\Omega = \sqrt{\xi} \, v_F/a \quad , \quad \Gamma = \Gamma_b + \left(v_F/a\right) ,$$ (14)

where the Fermi velocity v_F = 1.29 X 10^8 cm/s and Γ_h = 2.5 X 10^{13} s^{-1} in silver. The plasma frequency is related to the density of the electrons in our case, $\omega_p^2 = 4\pi ne^2 \backslash m_c \approx 4.0$ X 10^{31} s^{-2}. The electric fields are scaled to the parameter χ_3, which we assume to be real and positive.

The bistable behavior arises from the nonlinear relation between the local field inside the metal particle E_L and the propagating field E_0. The condition on tuning of the laser frequency to observe bistability is that

121

$$Re\left(2\epsilon_d + \epsilon_{mi}\right) < 0$$

and

$$\left|Re\left(2\epsilon_d + \epsilon_{mi}\right)\right| > 3\left|Im\left(2\epsilon_d + \epsilon_{mi}\right)\right|. \tag{15}$$

The first inequality is based on the assumption that $\chi_3 > 0$. The second inequality is violated near the resonance frequency Ω and we do not find optical bistability for this regime. Furthermore, the bistable characteristics for a laser tuned near this frequency would be very sensitive to tiny fluctuations in size and shape of the particles. The curve of the local field intensity versus the propagating field intensity is shown in Fig. 3 for particles of size $a = 5$ nm at a wavelength of 500 nm. At this wavelength the penetration depth of the electric field is about eight times larger than the particle radius $2\pi a \sqrt{Re(\epsilon_{mi})}/\lambda \approx 0.13$ and our use of the quasistatic approximation is justified.

Equation (11) is integrated using a forward-difference scheme. At each new point at which the propagating-electric field is found, we require that the local-field intensity $|E_L|^2$ be determined from Eq. (2) together with Eq. (3); the result is a cubic equation. In the regions of propagating-field intensity where three real roots are found, the solution chosen depends on the intensity of the input electromagnetic field and its previous history. If the input intensity is ramped up sufficiently slowly from zero, then the steady-state solution on the lower branch of the local-field intensity is chosen and it remains on this branch until the turning point marked A is reached in Fig. 3. Along the lower branch, the medium is highly dispersive and $|Re(2\epsilon_d + \epsilon_m)| \gg Im(2\epsilon_d + \epsilon_m)$. The absorption of the propagating-electric field in the medium is small for solutions along this branch. Six solutions that exemplify this behavior are shown in Fig. 4. They are labeled 1-6 and curve 6 is close to the turning point A in Fig. 3.

For input fields ramped to higher values of the intensity than turning point A in Fig. 3, the local-field intensity in the metal becomes large. Now the medium is driven into resonance by the local field and the medium is highly absorbing. The local field in the particles remains on the high branch until it reaches the turning point marked B in Fig. 3. The local field jumps discontinuously, but the propagating field remains continuous in the medium.

It should be noted that this switch in local fields occurs inside the medium, say

at length L and that at L , the effective-medium dielectric function undergoes a discontinuity. This boundary separating the high and low local field branches would scatter radiation in the backward direction, as does a similar boundary analyzed earlier for a different system[17], as well as for the nonlinear oscillator model[18]. However, for the concentrations in this paper, $f = 10^{-3}$, the discontinuity in ϵ^* is quite small.

The occurrence of the dielectric boundary is exhibited in Fig. 4 by curves 7-10. The knee in the curve separates a high-absorption regime from a low-absorption regime. As the input field is increased, the knee in the propagating intensity curve moves to the right; and as it is decreased, the knee moves to the left until it continuously moves back to the input and the local field is then on the lower branch throughout the entire medium.

We can study different effects of the internal boundary by observing the transmission characteristics for the silver composite cut to two different lengths. We show two examples in Figs. 5 and 6. In Fig. 5, the medium is short enough that the boundary does not appear in the medium. The transmitted intensity decreases when the input intensity reaches point A, but the output intensity is greater than that at point B (Fig. 3). Increasing the output intensity does not significantly change the absorption in the medium, so the transmission is again linear with the input field. As the input field is decreased, a change in slope occurs in the transmitted intensity and the output intensity remains nearly constant. The absorption in the medium is no longer constant; the knee in the intensity curve discussed in Fig. 4 is inside the medium. A further lowering of the input intensity will result in the motion of the boundary out of the medium at the input face and eventually the transmitted intensity will smoothly join the low-intensity solutions.

In Fig. 6, the medium has a length that is longer than can sustain the large local field, corresponding to the upper-branch solution, across the sample. The boundary now appears inside the medium. The contrast between high- and low-output intensities is greater and the nearly constant output intensity is sustained over a much larger range of input intensities.

6. CONCLUSIONS AND OBSERVATIONS

We find that the metal composite glasses can exhibit optical bistability when driven by a strong laser. The bistability is intrinsic since it does not require the use of an external cavity or other forms of optical feedback, and its features are analogous to the nonlinear oscillator model[18]. Our results are not restricted to photochromic glasses or metal colloids, but they could also be applicable to

inhomogeneous polymer solutions that have a larger χ_3 with metal particles dispersed in them to increase the effective nonlinearity.

We find that further tuning of the laser frequency toward the blue can give a significant decrease of the input intensities and the results presented here are not significantly altered for 20% variations of the particle sizes. In our studies the backward propagating wave is entirely negligible, and in higher particle concentrations the absorption may be too large to allow a significant reflected intensity from occurring at a boundary inside the medium.

The response time of these materials is determined by the relaxation time of the electrons excited by the applied field. According to Hache, Ricard, and Flytzanis[3], χ_3 can be enhanced in small particles by quantum size effects on the free electrons. This is due to a breakup of the continuum into a quasicontinuum. Therefore, the response time is of order Γ^{-1}; this is less than 1 ps for silver. For silver, they also calculate χ_3 (Ag) = 2.4 X 10^{-9} esu. With this value we estimate the intensity levels for the bistable switching to be around 100 MW/cm^2 at a wavelength of 450 nm.

REFERENCES

1. J. C. Gardland and D. B. Turner, eds., *Electrical Transport and Optical Properties of Inhomogeneous Materials*, AIP Conf. Proc. 40 (1978); D. Richard, "Nonlinear optics at surfaces and in composite materials," in *Nonlinear Optics: Materials and Devices*, C. Flytzanis and J. L. Oudar, eds. (Springer-Verlag, Berlin, 1986), p. 154.

2. D. Ricard, Ph. Roussignol, and Chr. Flytzanis, Opt. Lett. 10, 511 (1985).

3. F. Hache, D. Ricard, and C. Flytzanis, J. Opt. Soc. Am. B3, 1647 (1986).

4. K. M. Leung, Phys. Rev. A33, 2461 (1986); C. M. Bowden, R. Inguva, J. W. Haus, and N. Kalyaniwalla, Opt. News 13(9), 116 (1987).

5. J. W. Haus, N. Kalyaniwalla, R. Inguva, and C. M. Bowden, J. appl. Phys., 65, 1420 (1989).

6. R. K. Chang and T. E. Furtak, *Surface Enhanced Raman Scattering* (Plenum, New York, 1982).

7. J. W. Haus, R. Inguva, and C. M. Bowden, "Effective medium theory of nonlinear ellipsoidal composites," submitted to Phys. Rev. A41, 1670 (1990).

8. M. J. Bloemer, M. C. Buncick, R. J. Warmack, and T. L. Ferrell, J. Opt. Soc. Am. B5, 2552 (1988).

9. L. K. H. Van Beek, "Dielectric behavior of heterogeneous systems," in *Progress in Dielectrics*, J. B. Birks, ed. (CRC, Cleveland, OH, 1967), Vol. 7, p. 69.

10. D. Stroud, Phys. Rev. B12, 3368 (1975).

11. G. S. Agarwal and R. Inguva, Phys. Rev. B30, 6108 (1984).

12. D. Stroud and P. M. Hui, Phys. Rev. B37, 8719 (1988).

13. A. Yariv and P. Yeh, *Optical Waves in Crystals*, (Wiley, New York, 1984), p. 563.

14. J. Hajto and I. Janossy, Philos. Mag. B47, 346 (1983); M. Dagenais and W. F. Sharfin, Appl. Phys. Lett. 45, 210 (1984); D. A. B. Miller, A. C. Gossard, and W. Wiegman, Opt. Lett. 9, 162 (1984).

15. C. M. Bowden and C. C. Sung, Phys. Rev. A19, 2393 (1979); F. A. Hopf and C. M. Bowden, Phys. Rev. A32, 268 (1985); J. W. Haus, C. C. Sung, C. M. Bowden, and J. M. Cook, J. Opt. Soc. Am. B2, 1920 (1985). Y. Ben-Aryeh, C. M. Bowden, and J. C. Englund, Phys. Rev. A34, 3917 (1986).

16. D. S. Chemla and D. A. B. Miller, Opt. Lett. 11, 522 (1986).

17. Y. Ben-Aryeh, C. M. Bowden, and J. C. Englund, Opt. Comm. 61, 147 (1987).

18. J. W. Haus, S. Wang, M. Scalora, and C. M. Bowden, Phys. Rev. A38, 4043 (1988).
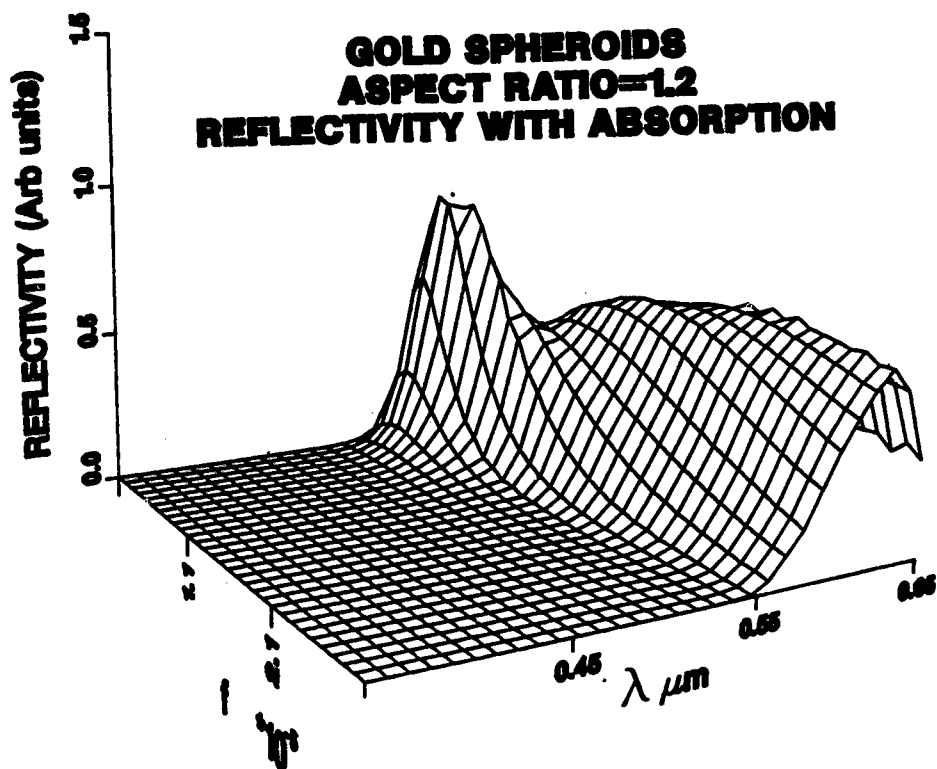
Fig. 1. Conjugate reflectivity for spheroids plotted as the concentration and wavelength is varied; use Eq. (28).
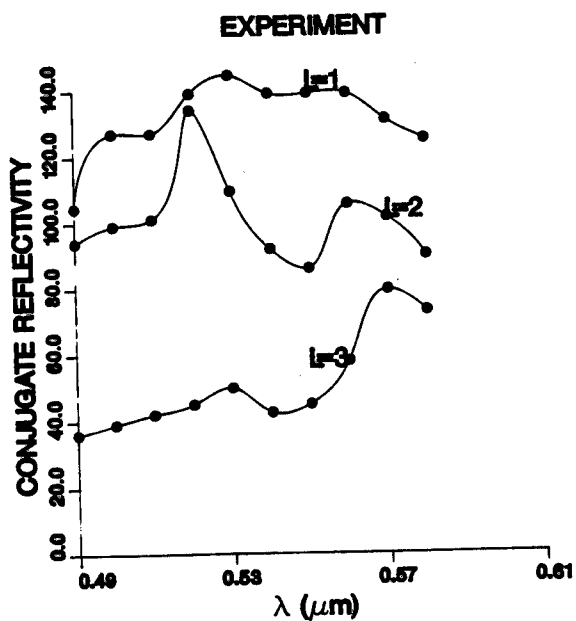
Fig. 2. Conjugate reflectivity data for the RG6 glasses. The curves are a spline fit to the data and are not from the theory.
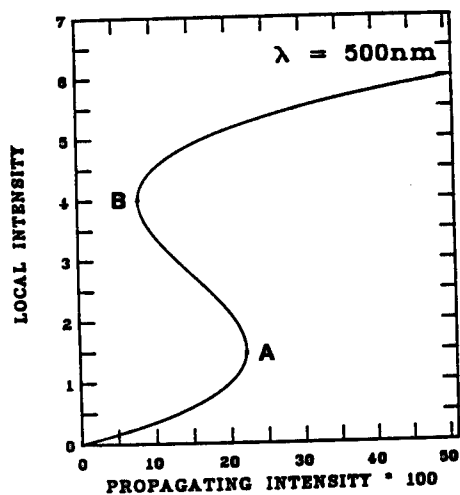
FIG. 3. Local-field intensity vs propagating intensity from Eq. (8). The particle radius is 5 nm.

126
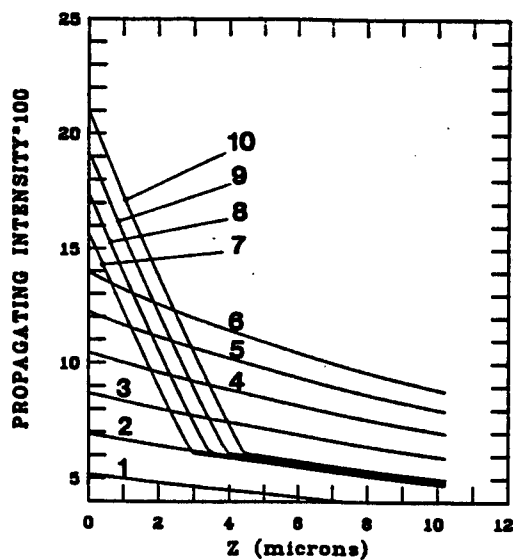
FIG. 4 Intensity of light in the medium. The input intensities are regularly spaced.
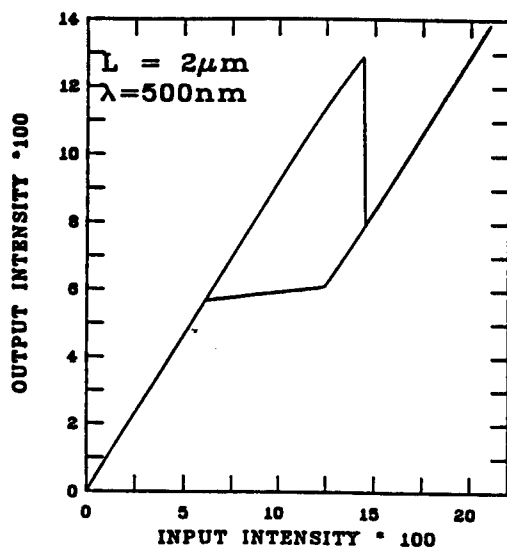
FIG. 5 Input vs output intensities for a 2-μm-thick sample. The particle radius is 5 nm.
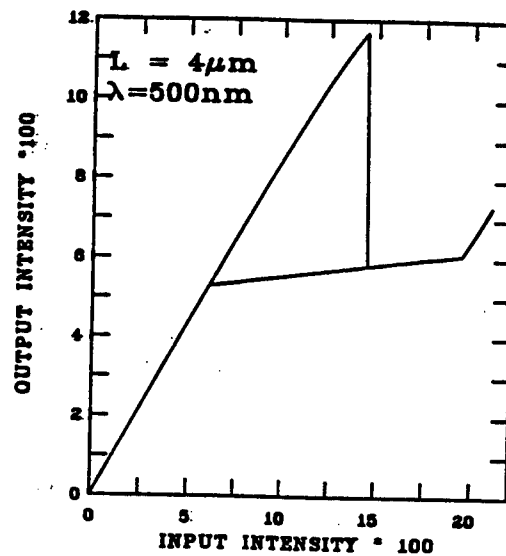
FIG. 6 Same as in Fig. 5 with a 4-μm-thick sample.

Reaction Kinetic Model (PANDORA) with Reaction Induction Times
for Molecular Collisions (U)

Dr. A.J. BRACUTI and Mr. D.S. CHIU
US Army ARDEC, Picatinny Arsenal, NJ 07806-5000

and

Dr. A.K. MACPHERSON
Packard Laboratory #19
Lehigh University, Bethlehem, PA 18015

ABSTRACT

The traditional method for solving chemical reaction-rate equations requires solving a stiff set of ordinary differential equations where the chemistry is assumed to follow Arrhenius behavior. The approach in this work also assumes Arrhenius chemistry, but in addition includes the effect of reaction induction periods which is controlled by the probability of sufficient collisions occurring during the time steps involved. This is particularly important when three-body collisions are being considered, although the probability of such collisions may be small. The inclusion of these induction periods requires a different numerical scheme than is ordinarily used. An additional feature of this solution scheme which is not usually found in traditional ones is the enforcement of strict conservation of atoms throughout the calculation. Furthermore, this program has been designed to be a user-friendly code which can be run on either a personal computer or workstation.

The use of this kinetic code (PANDORA) is demonstrated by modeling the oxidation of hydrogen; and, the initial combustion phase of a liquid within a liquid propellant combustor with a two-dimensional hydrodynamic/chemistry code (VULCAN) developed by the authors. Since VULCAN can utilize either equilibrium chemistry or reaction kinetic data, the sample problems in this study, have been addressed using only kinetic data from PANDORA.

INTRODUCTION

Traditionally, reaction kinetics computer programs used to solve combustion problems have been designed to run on large main frame computers. Running these programs usually required considerable study of instruction manuals and careful input of the necessary data in a specific format. More importantly, these codes also need a reaction mechanism or a weighted reaction scheme which would guide the reaction to completion. The intention of this study was to develop a user friendly program which required neither reaction mechanisms nor weighted reaction schemes and was similar in format and ease of operation to the chemical equilibrium program (MCVECE) developed by the authors[1].

A prototype code named PANDORA was produced which both prompts the user for necessary input and contains a chemical data library from which the appropriate reactions and rate constants can be extracted automatically[2].

In the initial version of PANDORA, Arrhenius chemistry was assumed and a semi-implicit time-integration scheme was used to solve the differential equations. The integration time-steps for each differential equation in this scheme are based on characteristic times which are deduced from the time required to reduce any component in the reaction to zero.

Closer examination of the integration time-steps involved in this solution revealed that often there was insufficient time for enough molecular collisions to occur for the reaction to proceed because of the magnitude of various Arrhenius parameters. As a result, longer computation times were required for reactions to reach equilibrium than anticipated. For example, reactions involved in the dissociation of hydrogen require integration time-steps of the same order of magnitude needed for an electron to move one Angstrom. In order to model this physically unrealistic situation, the concept of an induction time for each reaction was introduced. This is particularly important for three-body collisions where the induction times often will be quite long. Using simple kinetic theory, the times required for all atoms involved in each reaction to collide were calculated and an induction period was introduced. If the induction period was longer than the characteristic reaction time, the reaction was not considered to start until the induction time period had elapsed. The Arrhenius parameters were then used to calculate the development of the reaction. This method should improve the total computation time required for the reaction to reach equilibrium.

The induction times were calculated using simple kinetic theory and cross sections from Lennard-Jones potentials. If the reaction cross-sections for the colliding particles are X_i and X_j, the number densities are N_i and N_j, the molecular masses are M_i and M_j, and the temperature is T, then the time t_{ij} for all particles to collide once is given by

$$t_{ij} = \frac{1}{2N_i X_{ij}^2 \sqrt{2\pi k T M_{ij}}}$$

(1.1)

where

$$X_{ij} = \frac{1}{2}(X_i + X_j)$$

(1.2)

$$M_{ij} = \frac{1}{M_i} + \frac{1}{M_j}$$

(1.3)

In order to utilize this induction period concept effectively, it was necessary to use a solution scheme which did not depend upon the previous time derivatives of species concentrations. The semi-implicit time-advance scheme mentioned above and fully described by A. Macpherson, A. Bracuti and D. Chiu[2] satisfied this requirement.

In the present work the collision times were calculated assuming atoms have Lennard-Jones collision cross-sections. Other values for atomic radii could also be chosen if desired.

Also, in traditional solution methods it is not possible to actually conserve atoms due to round-off errors that result from the discrete integration time-steps. Using the present semi-implicit time integration scheme, a limit may be set for the maximum allowable error on all atoms in order to enforce a strict conservation of atoms throughout the calculation. The method of calculation is described in the appendix.

Combustion problems were addressed assuming both instant chemical equilibrium and time dependent chemistry. First the combustion of hydrogen in two different pressure/temperature regimes was examined in an effort to demonstrate the utility of the PANDORA code with a relatively simple reaction.

Finally, the combustion of a flowing fluid in a liquid combustor was addressed. The combustor employed was a simple converging combustion chamber with a moving piston controlling liquid input at one end and closed at the other end. This problem is of general interest and is particularly important in regenerative liquid propellant guns. In this combustor, described more fully in the next section, piston motion forces liquid into the chamber where a wave is generated in the reacting liquid.

Normally, it is assumed that chemical equilibrium is reached instantaneously and should exist throughout the combustion chamber. If for a given system this assumption is not valid, then reaction kinetics rather than equilibrium chemistry must be employed to describe the combustion process.

A more complex case which couples hydrodynamics with chemistry was modeled using the hydrodynamic code named VULCAN[3] which can employ either the chemical equilibrium or reaction kinetic approaches. Results obtained from VULCAN models with reaction kinetic data (PANDORA) for the initial combustion phase in a liquid combustor will be presented here.

RESULTS AND DISCUSSION

If a kinetic code truly simulates a reaction mechanism, then it should also model chemical equilibrium after sufficient reaction time has elapsed. After a suitably long reaction time, a kinetic code should yield the same results as an equilibrium code. If the results from the two codes disagree, then the reaction mechanism utilized by the kinetic code is incorrect. In order to correct this in most kinetic codes, the reaction path or mechanism must be adjusted until agreement is achieved with the equilibrium code.

PANDORA, however, differs from other kinetic codes because it does not require a predetermined reaction mechanism. Instead it requires menus of chemical species and reactions from which it will choose a set of kinetic equations based on the input reactant chemical formulas. The total number of reaction equation utilized by the code may be very large and long computation time is required. In order to make the code to run more efficiently and faster, molecular collision theory is used to prioritize the reaction equations and thus arbitrarily set a reaction scheme. For example using the reaction cross sections, it is possible to estimate from kinetic theory the time for reacting species to undergo the necessary number of collisions in order to complete the partial reaction step. Thus for a given integration time step, only certain reactions will satisfy this criterion and all others will be neglected. At the next time step, some of the neglected reactions will be included as sufficient time will have elapsed for these to have undergone the required number of collisions. Therefore, additional species and reactions can be supplied to PANDORA's data library until agreement is reached with the equilibrium code.

Thus, comparison of data obtained from both PANDORA and MCVECE (or other equilibrium code) will reveal when equilibrium is attained. This is particularly important in the interior ballistic cycle of a gun where the propellant combustion period is quite short. Normally, it is assumed that chemical equilibrium is reached instantaneously and should exist throughout the combustion chamber. If the time required to reach equilibrium is of the same order of magnitude as the interior ballistic combustion cycle time, then reaction kinetic must be employed to realistically model it over time and space.

The initial reaction studied was the oxidation of 1 mole of H_2 with 1/2 mole of O_2 at both 1000 $^\circ$K / 6 atm and 3600 $^\circ$K / 20 atm.

$$H_2 + \frac{1}{2}O_2 \text{----> } H_2O \tag{1.4}$$

The combustion products obtained in a given period of time (0.01 sec) with both the kinetic and the equilibrium chemistry codes are compared in Table 1. Both NASA-Lewis[4], MCVECE and PANDORA yield similar results at both conditions. This suggests that PANDORA chose a reasonable set of reactions; and that equilibrium was attained in the specified reaction time. An output from

PANDORA for the oxidation of H_2 at 3600 $^\circ$K / 20 atm presented in Figure 1 indicates that OH and H are initially formed from H_2 and O_2 before 10^{-18} sec has transpired but little apparent interaction among the species occur during the time interval between 10^{-18} and 10^{-10} sec. After 10^{-10} sec have elapsed, H_2O starts to form and equilibrium is reached after 10^{-3} sec. It must be noted, however, that the published H_2O/OH reaction rates varied greatly and mean values were used in this work. It was also found that in this simple case, the equilibrium constants play an important role in determining the final products in the PANDORA code.

The thermal decompositions of neat hydroxyl ammonium nitrate (HAN, NH_3OHNO_3), neat triethanol ammonium nitrate(TEAN, $(C_2H_5O)_3NHNO_3$) and a solution of 60% HAN, 20% TEAN and 20% water were also modeled with PANDORA and compared with equilibrium results obtained from the NASA-Lewis and MCVECE codes. Since gaseous species are required by kinetic codes, initial decomposition products used in PANDORA for both HAN and TEAN were those reported in the literature[7]. The initial breakdown equation for the HAN used was

$$2NH_3OHNO_3 \approx N_2O + 2H_2O + O_2 + HNO_3 + NH_3 \tag{1.5}$$

For HAN, combustion products in a given period of time are shown in Table 2. A comparison with the equilibrium NASA-Lewis and MCVECE codes shows good agreement at a density of 0.2 g/cc. This indicates that the initial breakdown equation for HAN is valid. The initial breakdown equation for the TEAN used was

$$(C_2H_5O)_3NHNO_3 \approx 3C_2H_4O + NH_3 + HNO_3 \tag{1.6}$$
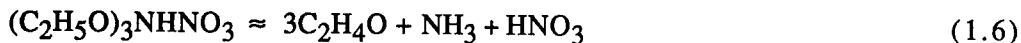
For TEAN at a density of 0.2 g/cc, the results are shown in Table 3. The PANDORA code predicts more water, methane and carbon monoxide than does the equilibrium code. This suggests that a different reaction mechanism may be involved. It may require more reaction equations in the chemical library for the PANDORA.

The results obtained with PANDORA, NASA-LEWIS and MCVECE for the thermal reaction of a solution of 60% HAN, 20% TEAN and 20% water at a density of 0.2 g/cc are presented in Table 4. The PANDORA, NASA-Lewis and MCVECE codes are in good agreement with water, nitrogen and carbon dioxide predicted as the main combustion products. The small observed differences between the two codes are within computational error considering the number of equations involved and time steps used.

Combustor Modeling

The combustor configuration shown in Figure 2 uses a shower head-like piston injector which mechanically atomizes the liquid propellant into a cylindrical combustion chamber which is 18 cm in length and 22 cm in diameter. Only the initial combustion phase of a liquid propellant is studied here. All calculations were carried out on a square grid of 0.5 cm x 0.5 cm. This two-dimensional model simulates a one-dimensional case if one assumes that the combustion products are uniform in the y-direction.

In this simulation, at time zero the piston is located at the end of the combustion chamber with liquid propellant behind it and a bullet at the front. At initiation, the liquid propellant droplets are injected into the combustion chamber immediately after the piston starts to move backward away from the venting aperture. If the equilibrium code is used, the reactions are assumed to be instantaneous and the piston velocity is input as 1000 cm/sec. With the kinetic code, however, this assumption is unnecessary because the piston velocity can be calculated from the fluid pressure. During this study, the piston is allowed to travel backward two grid points and the simulation terminates when the bullet starts to move.

Combustion modeling of the beginning of the combustion stoke was undertaken using a nominal LP1846 formulation containing 60% HAN, 20% TEAN and 20% water. This liquid propellant was used both as igniter fluid with a loading density of 0.02 g/cc and as propelling charge with a loading density of 1.4 g/cc. The initial temperature for both liquid propellant and igniter was 2600 °K. In addition, the shot start pressure was assumed to be 20 MPa (2900 psi). Instantaneous burning of the liquid propellant and no heat loss between the combustion gases and the combustor walls were assumed.

With this model, it was assumed that 15% of injected liquid propellant was in the gaseous phase while 85% was assumed to be liquid droplets distributed over four grid points in front of the piston. The burning rate was assumed to be 0.02 m/s. The averaged density, pressure and temperature at varies times at which the pressure pulse has just reached the piston are presented in Figure 3, 4 and 5, respectively. Pressure near the piston face as a function of time as shown in Figure 6 indicates that the pressure generated is unstable. At all times during the simulation, high pressure and density regions are observe near the piston while the low temperature regions are observe consistently before the piston. The data suggest that equilibrium is not achieved during this simulation.

CONCLUSION

The sample calculations have shown that there is good agreement between the kinetic code and chemical equilibrium code provided sufficient time is allowed to reach thermodynamic equilibrium and all the reaction equations are included.

The addition of induction time periods to PANDORA have reduced the modeling computation time by a factor of 50% in most cases when compared to the same reactions modeled with the initial version of PANDORA without the molecular collision theory modification. These results are not necessarily representative of the optimum improvement expected, since only limited data have been obtained using molecular cross-sections from Lennard-Jones potentials exclusively. The computation time is sensitive to the choice of molecular cross-sections since they will control the induction times for the many difference reactions in the computation.

The results of the VULCAN/PANDORA liquid combustor simulation suggest that equilibrium condition should not necessarily be assumed during the initial combustion phase of the combustor.

REFERENCE

[1] T. Vladimiroff, Y.P. Carignan, D.S. Chiu and A.K. Macpherson, "The Development of a User Friendly Thermodynamic Code for the Personal Computer", 24 th JANNAF Combustion Meeting, Volume II, Pg. 39, 5-9 October 1987.

[2] A. Macpherson, A. Bracuti and D. Chiu, "Reaction Kinetics of HAN and Water Mixtures Using a Personal Computer", Technical Report ARAED-TR-89008, ARDEC, Picatinny Arsenal, New Jersey, Sept. 1989.

[3] A. Macpherson, A. Bracuti and D. Chiu, "Hydrodynamic Modeling of the LP Gun", 26th JANNAF Combustion Meeting, JPL, Pasadena, CA, Oct. 1989.

[4] S. Gordon and B.J. McBride, "Computer Program for Computation of Complex Chemical Equilibrium Compositions...", NASA SP-273, 1973.

[5] M.R. Laydi, J.C. Miellou and J.P. Parisot, "Some Computational Techniques for Studying Chemical Kinetics" in Numerical Simulation of Combustion Phenomena, Ed R. Glowinski, B. Larrouturou and R. Temam Springer-Verlag 1985 Berlin.

[6] S.H. Lam and D.A. Goussis, 12th IMACS World Congress on Scientific Computation Paris, 1988.

[7] J.T. Cronin and T.B. Brill J. Phys. Chem. 90, 178, 1986.

Table 1. Comparison of chemical equilibrium and reaction kinetic results for $H_2 + \frac{1}{2} O_2$ at 1000 $^\circ$K / 6 atm and 3600 $^\circ$K / 20 atm.

Mole Fractions

	1000 $^\circ$K / 6 atm			3600 $^\circ$K / 20 atm		
Species	NASA	MCVECE	PANDORA	NASA	MCVECE	PANDORA
H_2O	1.000	1.000	.9907	0.58	0.56	0.56
H_2	<1.E-5	<1.E-5	.0015	0.15	0.16	0.18
H	<1.E-5	<1.E-5	.0020	0.06	0.07	0.05
OH	<1.E-5	<1.E-5	.0040	0.12	0.13	0.12
O	<1.E-5	<1.E-5	.0004	0.03	0.03	0.02
O_2	<1.E-5	<1.E-5	.0014	0.05	0.06	0.06

Table 2. Comparison of chemical equilibrium and reaction kinetic results of HAN, NH_3OHNO_3, at a density of 0.2 g/cc.

Mole Fractions

Species	NASA-Lewis	MCVECE	PANDORA
H_2O	0.5000	0.4999	0.4997
O_2	0.2496	0.2496	0.2510
N_2	0.2497	0.2497	0.2490
NO_2	0.0001	0.0001	0.0002
NO	0.0006	0.0006	2.13E-5
OH	0.0002	0.0001	3.59E-6

Table 3. Comparison of chemical equilibrium and reaction kinetic results of TEAN, $(C_2H_5O)_3NHNO_3$, at a density of 0.2 g/cc.

Mole Fractions

Species	NASA-Lewis	MCVECE	PANDORA
H_2O	0.198	0.200	0.225
N_2	0.091	0.091	0.100
CH_4	0.188	0.188	0.273
H_2	0.157	0.155	0.047
CO	0.148	0.147	0.297
CO_2	0.103	0.102	0.048
NH_3	0.002	0.002	<1.E-5
NH_2	<1.E-5	<1.E-5	0.007
NHO	<1.E-5	<1.E-5	<1.E-5

Table 4. Comparison of chemical equilibrium and reaction kinetic results of a solution 60% HAN, 20% TEAN and 20% water at a density of 0.2 g/cc.

Mole Fractions

Species	NASA-Lewis	MCVECE	PANDORA
H_2O	0.6991	0.6990	0.6767
N_2	0.1633	0.1633	0.1471
CO_2	0.1203	0.1201	0.1319
CO	0.0082	0.0084	0.0016
H_2	0.0082	0.0082	0.0018
OH	0.0006	0.0007	0.0027
NO	0.0002	0.0002	<1.E-5
O_2	0.0001	0.0001	0.0041
NO_2	<1.E-5	<1.E-5	4.73E-4

137

APPENDIX

METHOD OF CALCULATION

THE PANDORA CODE

The equations required for the chemical kinetics of the fluid system will involve variable temperature and density. The present calculations were undertaken at constant density as this is most easily compared with the minimum free energy calculations. The changes required to allow variable density are minimal. The basic equation for the change of the concentration c of the ith species is

$$\frac{dc_i}{dt} = \Sigma b_{ij} r_j \qquad (2.1)$$

where b_{ij} is the ith species in the jth equation and r_j is defined as the net rate by the mass-action equation

$$r_j = k_f \Pi c_i - k_r \Pi c_i \qquad (2.2)$$

where the forward reaction rate k_f is related to the reverse reaction rate k_r by the equilibrium constant K_{eq}

$$\frac{k_f}{k_r} = K_{eq} \qquad (2.3)$$

The equilibrium constant used expresses the ratio between the number of moles of products and the number of moles of reactants. In the case where the number of moles of products and reactants are the same, then $K_{eq} = K_p$ the usual equilibrium constant. In the case that

$$\Delta N = \Sigma N_p - \Sigma N_r \qquad (2.4)$$

is not zero then

$$K_{eq} = K_p (RT)^{-\Delta N} \qquad (2.5)$$

In the case studied here the pressure P will rise according to the relation for ideal gases

$$\frac{dP}{dt} = RT\Sigma \frac{dc_i}{dt} \tag{2.6}$$

Equation 2.1 is used to integrate the change in concentration in time. Conceptually the integration of this derivative is quite straight forward, but the real difficulties are imposed by the limitation in accuracy of the computer. The first problem is the fact that the integration scheme may become unstable with time steps which are too large. This can be overcome by various methods. The most common method involves calculating the high order derivatives, for example the first four derivatives and enforcing continuity in all derivatives. A major problem with this method is that sudden changes cannot be made in a variable. These changes must be undertaken gradually. As will be seen this is satisfactory on a main frame computer where calculation time is not the chief concern. When solving the problem on a personal computer with reasonable run time, it is necessary to be able to make approximations which require sudden changes in the function in order to conserve certain quantities e.g. atom numbers. Thus a semi-implicit time advance scheme[5] was chosen.

If (2.1) is written as

$$\frac{dx_i}{dt} = P_i - x_i L_i \tag{2.7}$$

where P_i are the production terms and L_i are the loss terms for the species x_i. The usual finite difference form of the semi-implicit scheme is

$$\frac{x_i(j+1) - x_i(j)}{\tau} = P_i(j) - x_i(j+1)L_i(j) \tag{2.8}$$

where the right hand side can be written as $\left[\dfrac{1}{1+\tau L_i(j)}\right]f(j)$. Now (2.7) can be

approximated by

$$\frac{dx}{dt} = P_i(\tau_j) - x_i(\tau)L_i(\tau_j) \tag{2.9}$$

integrating (2.9) we obtain

$$\frac{x_i(j+1) - x_i(j)}{\tau} = \left[\frac{1 - e^{-\tau L_i}}{\tau L_i(j)}\right]f(j) \tag{2.10}$$

where τ is the time step. This scheme is very stable and the original papers[5,6] should be consulted for details.

Equation (2.9) assumes that the equations are linear in x_i. There are equations where higher order equations occur. In the set of equations used for the present study, the highest order is two for which equation (2.10) becomes

$$2\tau\sqrt{L_i}\sqrt{P_i} = ln\left|\frac{x_j\sqrt{L_i} - \sqrt{P_i}}{x_j\sqrt{L_i} + \sqrt{P_i}}\right| - ln\left|\frac{x_{j+1}\sqrt{L_i} - \sqrt{P_i}}{x_{j+1}\sqrt{L_i} + \sqrt{P_i}}\right| \qquad (2.11)$$

The value of $x_{j+1} - x_j$ may be obtained from equation (2.11) although care must be taken in evaluating the absolute values of the log expressions. A characteristic time T_c is assigned to each equation such that one compound involved in the reaction will go to zero in T_c. Thus the time step is chosen from the minimum T_c for the reactions involved. It sometimes occurs that only a small number of reaction will be involved for an extended time with small time steps. An equilibrium assumption is invoked whereby it is assumed that the equations involved are considered to be in quasi-stable equilibrium and eliminated for one time step at the next largest T_c. This moves the system from the quasi-stable state and lets the reaction continue. The system would of course in time move out of the quasi-stable state but this assumption increase the speed of the calculation. The time which elapses before the quasi equilibrium state is assumed can be a user specified variable.
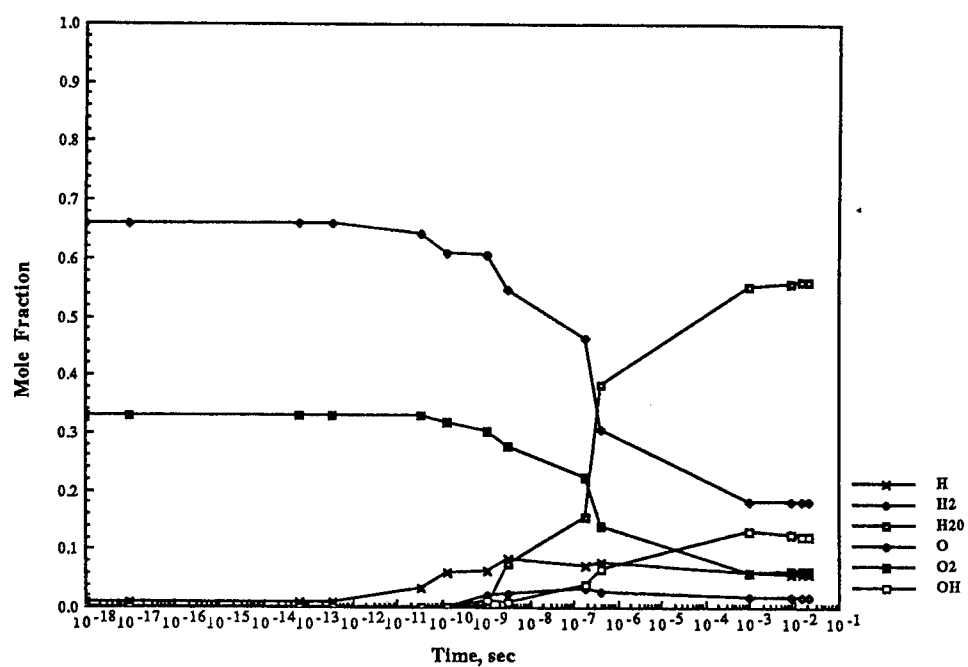
Figure 1.   Sample output from PANDORA of $H_2 + \frac{1}{2} O_2$ at 3600 °K / 20 atm.
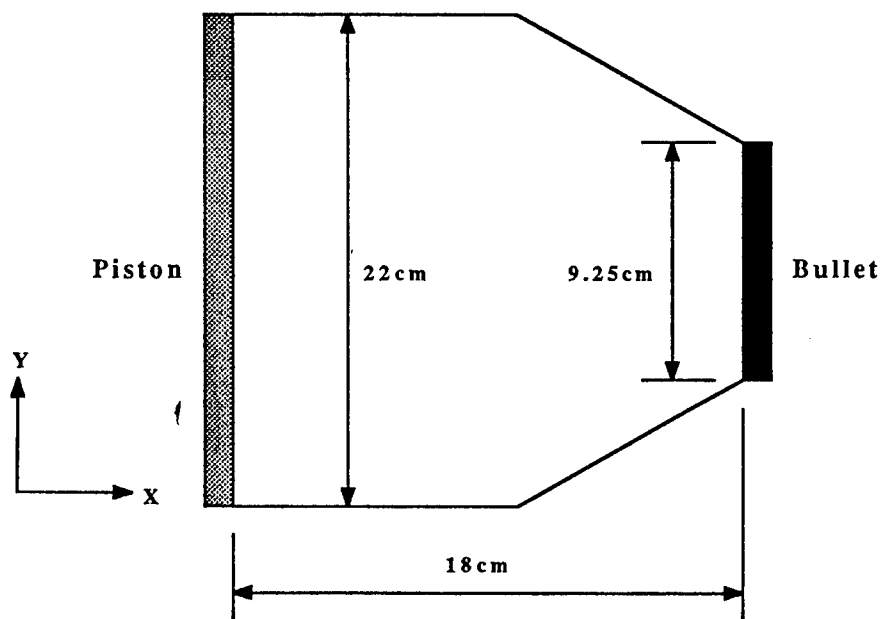
Figure 2.   Schematic diagram of the combustor.
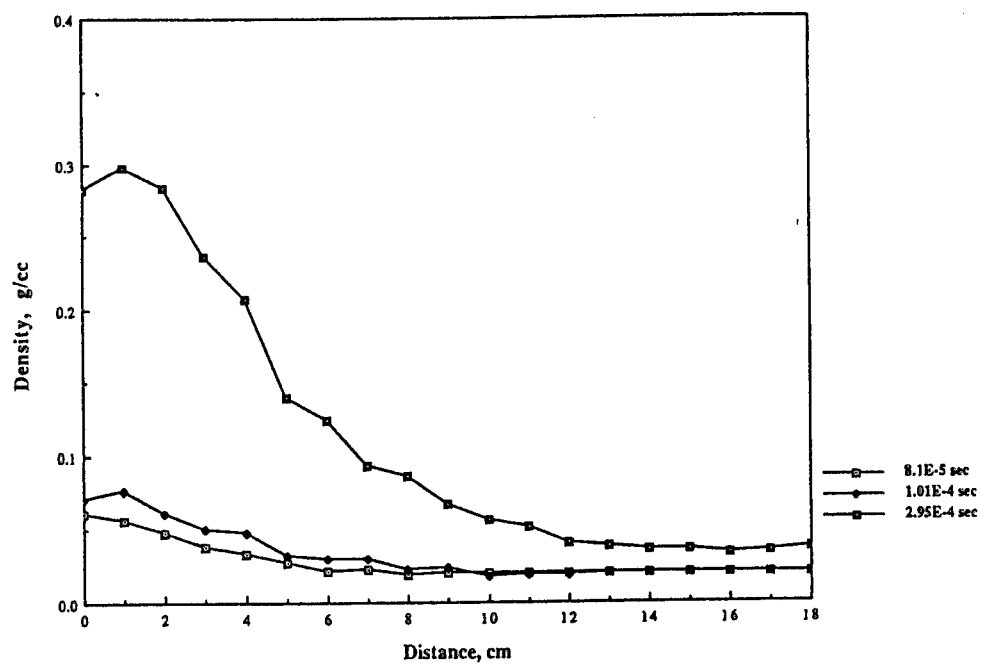
Figure 3.   Density variations along the combustor at various time.
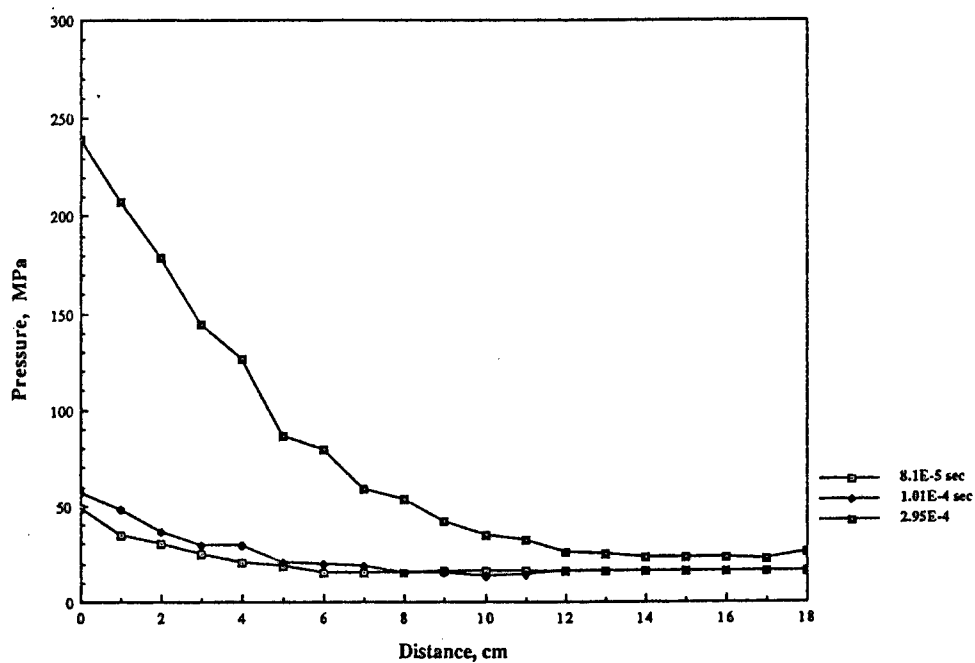
Figure 4.   Pressure variations along the combustor at various time.
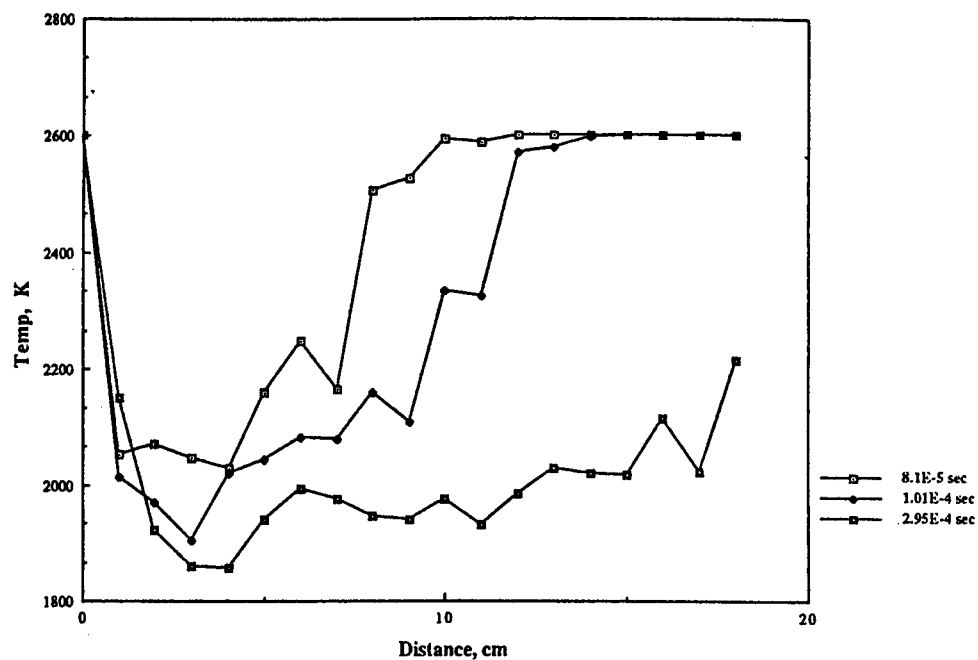
Figure 5.   Temperature variations along the combustor at various time.
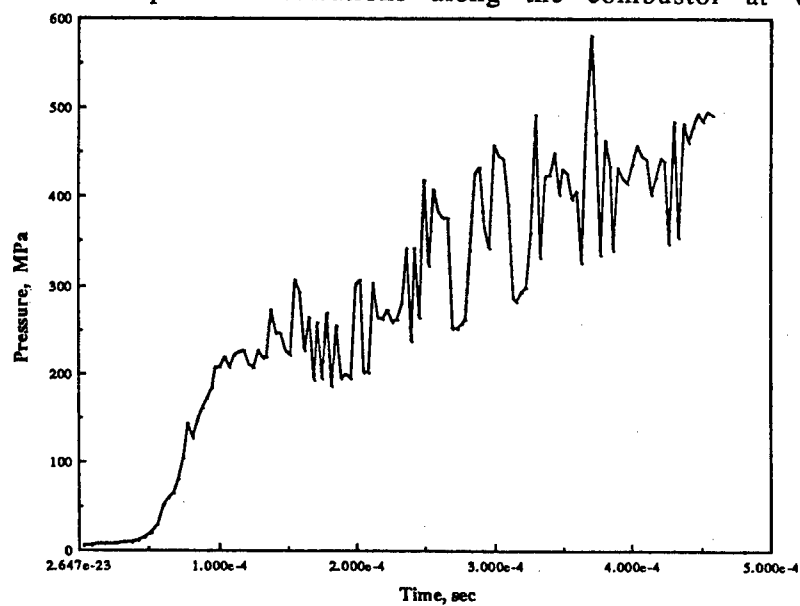
Figure 6.   Pressure near piston face.

MCT Solid-State Switching:
Revolution in Power Conditioning (U)

* Christopher G. Braun, Dr., CPT,
Robert Pastore, Maurice Weiner, Dr.
U.S. Army Electronics Technology and Devices Laboratory
LABCOM, Fort Monmouth, New Jersey 07703-5000

I. Introduction

This paper presents an overview of our work on MOS Controlled Thyristors (MCTs) and the major issues which are related to use of high-power switches. The research and development work on MCTs is supported by a joint collaboration of the Army, Navy, Air Force, NASA and industry. Many different approaches for device development and application demonstrations are taken by the various agencies to best meet their goals. Our program, representing the Army is funded by the Strategic Defense Initiative Office to develop high-voltage MCTs with a voltage standoff of at least 2,500 volts and a capability to turn-off 200 A/cm^2 in a few microseconds. The other major MCT device development program is conducted by the Air Force which has developed a near commercial 1,000 volt MCT. Part of our mission here at the Pulse Power Center is to provide the expertise to help transition these research devices towards meeting the needs of the government. Consequently, we are very involved in device testing, modeling, and application insertion issues.

The military is in the forefront of the development and use of high electric powers for use in a wide range of applications from missiles to electric vehicles to ships to space platforms. The control and management of these high powers is a growing issue in existing and future systems. The requirements placed on these high-power systems are strict; the criteria of weight, size, power dissipation, reliability have yet to be met for a tactical system. However, we are nearing a turning point with the recent developments of high-power technology and expect it to make significant impact in future systems.

In high-power systems there are several basic components - energy storage devices (such as capacitors, inductors, batteries), loads, transformers, connectors and transmission lines, and switches. In all areas there is a great need to decrease size and weight for tactical use. Often the performance and design of the entire power system depends on the limits of the switch; typically the power system is designed to avoid or to minimize the limits of the switch. The first high-power switches have been plasma devices such as thyratrons, spark gaps, or vacuum tubes. While capable of switching high powers, these devices have intrinsic drawbacks such as requiring excessive and constant cathode/heater

145

power, warm-up times of several minutes, relatively short lifetimes, and a low repetition rate. The advance of high-power solid-state devices eliminates these drawbacks and addresses a wide need from low to very high powers.

When considering the tradeoffs in design of a high-power system, we must look at component and system limitations. Roughly speaking the maximum power density per unit weight per unit size comes when the voltages are on order of 3-5 thousand volts. At lower voltages we have a wide selection of excellent components, however this voltage level requires relatively high currents which demand extensive bussing systems and result in high resistive losses. At higher voltages we encounter a threshold on insulators/surface-voltage breakdown which is only solved by dramatic increases in volume (separation of components) and use of insulators. The power utilities encounter the same constraints in transmitting high powers over long distances. Their solution is to use extremely high voltages (several hundred kilovolts) and large transmission towers.

Below an approximate level of several hundred volts there is an excellent selection of switches such that switching is no longer a limitation. Above this limit, we have a few choices: FETs, BJTs, SCRs, GTOs, and recently the MCT. Shown in Table 1. is a summary of the available high-power solid-state switching devices and an indication of the strong and weak points.

Table 1.

Device	Strong areas	Weak areas
FET	Fast (~10 ns on/off), easy control, turn-on/off, commercial 1 kV devices.	Above ~500 volts, the on-state resistance is high causing a large forward voltage drop and high power loss.
BJT	Fast (10-100 ns on/off), turn-on/off, commercial 1 kV devices.	Has a large forward voltage drop (~10 V) causing high power losses. Large drive current.
SCR	Very high voltages and currents (<10 kV, >1kA average), low forward voltage drop ($\leq$ 5 V).	Only turn-on.
GTO	Very high voltages and currents (<8 kV, 1 kA average), low forward voltage drop ($\leq$ 5 V).	Can turn-on/off but requires large control circuitry, slow (200 μs) turn-off.
MCT	Fast (100ns on, 2 μs off), turn-on/off, ease of control, low forward voltage drop ($\leq$ 5 V), samples at 1kV, research device 2.5 kV.	Not commercially available yet, high cost.

When one single switching device cannot meet the system requirement for voltage hold-off or current rating, we can use multiple devices together in series/parallel arrays. While conceptually simple, in reality this poses great challenges. The major difficulties are

transient voltage and current sharing, non-uniformity of the devices, temperature-related instabilities, isolation of control voltages from floating device voltage, losses due to snubbers and other compensating elements, and failure modes. A single device is usually rated at 50% of its maximum voltage/current. A group of five devices in series or parallel is de-rated by at least 20%, depending on compensating snubber networks. Further assemblies of groups of devices may also need to be de-rated. For large arrays of devices this is a losing game!

From the switching point of view there are only several main sets of high-power applications such as motor controllers, circuit breakers, etc. Generally, the switch and other components are carefully chosen to optimize each application. The requirements placed on the switch are quite different from one of these applications to another. Section V discusses application issues and examines these requirements.

The impact of an advanced high-voltage MCT is many fold: the ability to completely optically isolate MCT control circuitry, availability of high voltage devices would slash the size of series stacks, advances in the MOS processes are leading to a much higher turn-off current density than other similar devices, ability to operate over a wide temperature range. This is the basic motivation for the joint developmental program of MCT devices - to sufficiently advance the state of the art of high-power solid-state switching so that the switch is no longer the major limiting factor.

II. MCT Basic Operation

The MCT is a new class of thyristor developed by V. Temple at the General Electric Center for Research and Development (GECRD).[1,2,3,4] Figure 1 shows a diagram of the cross section of one of the many gate fingers in the device. Essentially, the device is a PNPN thyristor structure with the addition of a high density of MOS FET gates on the anode for turn-on and off. These gates serve several functions. They eliminate most of the external control circuitry as found with a Gate Turn-Off thyristor (GTO). Additionally, because of the high density of these gate cells, currently about $10^5/cm^2$, the on and off currents are forced to flow in a much more uniform and controlled manner than other thyristors. This leads to a significant increase in maximum current density.
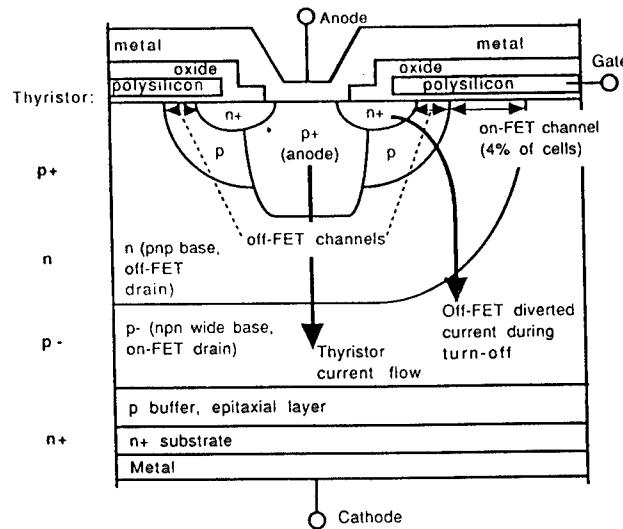
Figure 1. Cross-section of the unit cell for a MCT (not to scale). In the on-state, the current flows from the anode through the thyristor structure. Turn-off is initiated by diverting the current away from the upper PNP transistor by the off-FETs.

Only about 4% of the total cells are PMOS on-cells, the remaining are NMOS off-cells. In the off state, the N/P- junction is the reversed biased blocking junction. The upper P/N junction and the FET circuitry are less than 10 μm deep, the p- wide base is about 50 μm deep, and the substrate is several hundred μm thick. The cell repeat distance is 20 μm.

Turn-on of a MCT is caused by the on-cells injecting a small current to the base of the lower transistor (the NPN in Figure 1 and Figure 5). This causes the lower transistor to begin to conduct which draws charge from the base of the upper transistor. Current begins to flow though both transistors and positive feedback quickly latches the device on. In the on-state, the voltage drop is nearly that of a single forward biased P-N junction. Once the device is latched on, the on-cells are no longer needed and gated off.

The MCT is turned off by gating the off-cells on to shunt the entire current (10s-100 amps) around the upper transistor (the PNP in Figure 1 and Figure 5). This causes current to cease to flow in the upper transistor, this in turn shuts off the flow of current from the collector of the upper transistor to the base of the lower transistor. Consequently, the gain of the thyristor is reduced to less than one and the device begins to turn-off. The current fall time is determined by the carrier recombination time in the wide base of the lower transistor, typically a few microseconds During this time the off-cells must continue to divert current from the upper transistor otherwise the MCT will quickly revert to the on state.

III. MCT Testing

Here at the Pulse Power Center we are conducting a program to evaluate and fully characterize MCT devices. Since these devices are still in the development stage, very little data is available from the manufacture on the operational limits. We characterize these MCTs using various circuits with a wide range of resistive, capacitive, and inductive loads. Different gate driver circuits have been constructed and tested.

The MCT devices under test were made by GECRD and Harris Semiconductor. Most of these were samples of epitaxial-based MCTs which have a voltage rating of about 1,000 volts blocking and 600 volts operating. A key measurement is the Safe-Operating-Area (SOA). This delineates the range of voltages, currents, $\partial V/\partial t$, $\partial I/\partial t$, power, etc. where the device can be safely operated. While our measurements are by no means complete, we have verified that operation at 60% of the blocking voltage is possible without any compensating elements. Additionally, we have demonstrated that with a good snubber design these devices can be operated at 80% of blocking voltage.[5]
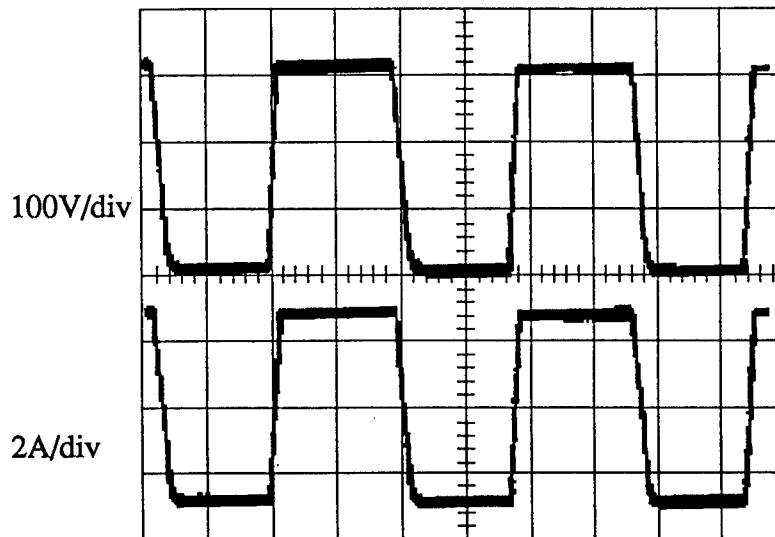
Figure 2. Current and voltage of a 30P50 MCT at a frequency of 50 kHz and 50% duty cycle. Applied voltage was 295 volts, peak turn-off current was 5.5 amps, average power switched was 826 watts. The MCT case temperature was 107 °C in a 30 °C oil bath.

Shown in Figure 2 is the current and voltage waveforms for high-frequency operation of a MCT device. This demonstrates that 50 kHz switching, as needed for high-frequency inverters, is possible with MCTs. We have tested a single device at 20 kHz (50% duty cycle) up to an average power of 2 kW - the highest continuous power demonstrated - at an efficiency of 97.6% (~50 W into the junction of the MCT).

149

There are several failure modes for MCTs. During turn-off current is shunted around the upper transistor through the off-FETs. Exceeding the maximum turn-off current results in local heating and damage near the off-FETs. If the rate of rise of the on-state current is faster than the current spreading time then the current will again be forced to conduct through a small region leading to heating and damage of the on-FETs. When the applied voltage is greater than the blocking voltage of the MCT, the blocking junction will collapse. This results in a discharge of internal device capacitance (if not any external source) that will most likely damage the thyristor. Unlike the fault tolerant high-power plasma switches we discussed in the introduction, these kinds of semiconductor devices can be easily damaged by exceeding the SOA.

Under the Army/SDIO contract GECRD is developing high-voltage diffusion-doped MCT devices. Recently, we tested the first sample 2,300 V (blocking voltage) device. Because the device was damaged rather early during testing, relatively little data have been obtained. We observed the first high-voltage device had a larger power dissipation than predicted. This particular device was not irradiated (which is the standard means to reduce the carrier lifetime in the wide base), so the long turn-off time was expected. However, we also measured a high forward voltage drop which was disappointing. Delivery and testing of more devices is necessary before further conclusions can be drawn.

Figure 3 shows the circuit diagram of a MCT series stack and Figure 4 shows the voltage division of this stack during one pulse.[5] The power supply feeds energy into a large capacitor bank (300μF) which is connected to the MCT stack. The gate drives of the MCTs are optically isolated and each gate drive is referenced to the anode voltage of its MCT. The total current and the voltages of each MCT is recorded. The temperature of the heat sink of the stack is also monitored via a fiber optic temperature probe. The MCTs were characterized individually and matched for uniform turn-on and off times. All MCT gates were pulsed at the same time.

The snubber network is used to help completely turn-on all devices at once by quickly dumping the capacitor energy into the MCT when it is first turned on. This prevents an overvoltage on a stage by a device that is slightly slower than others. The snubber is also used to absorb the initial energy during turn off. This removes stress from the MCT by momentarily diverting current to the capacitor and allowing the MCT time to turn off at low current.
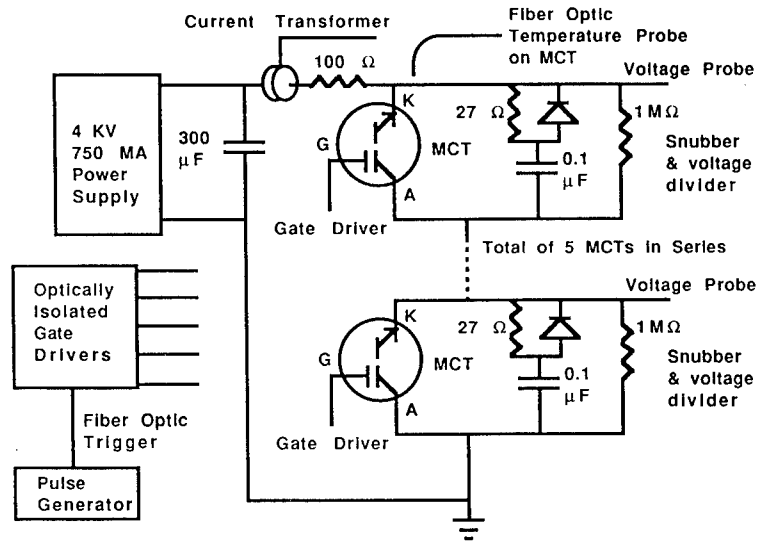
Figure 3. Circuit diagram of MCT series testing stack. The snubber network removes stress from the switch during the first part of the commutation process. The voltage divider ensures that the devices DC voltage share.
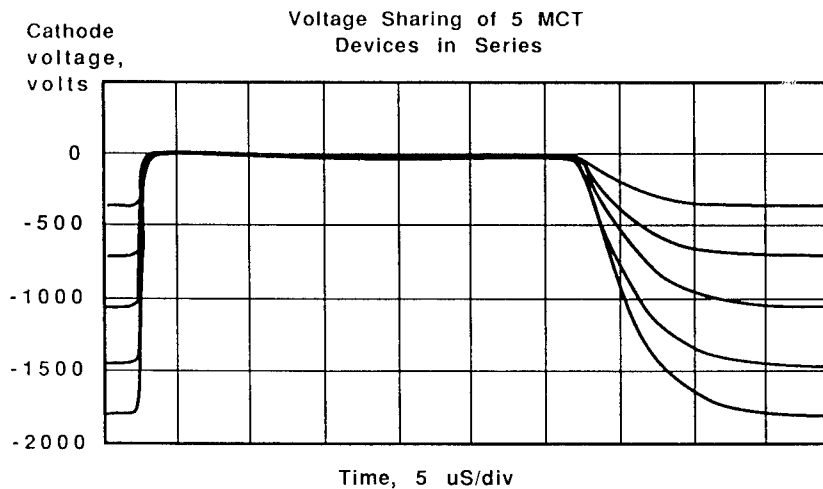
Figure 4. Shown is the voltage division of each stage of a series MCT stack using Harris 20P50 devices. The devices are initially holding off about 2,000 volts. The long turn-off time is due to the RC decay time of the snubber and not the MCT fall time.

151

The initial MCT devices have been tested over a wide temperature range from 72 K to 473 K. This data shows that MCTs are capable of operating over a very large temperature range without major reduction in capability.[6] If true for commercial MCTs this would be a boon to designers since the requirements for cooling would be greatly relaxed.

IV. Circuit Modeling of MCTs

Circuit-level models of power components like MCTs are vital since they give the system designer the tools to analyze system limits and tradeoffs. As a result, we are working to construct circuit-level models (SPICE) that accurately represent the MCT. The difficulty here is that the MCT is a complex device consisting of many thousand of MOS gates which regulate the flow of current through the thyristor structure. Further, we are still in the sampling stage of production; the device design continues to change at a rapid pace. As a result, the experimental data gathered is preliminary and not complete.

Temple et al. at GECRD have done considerable work in modeling MCTs. Using a finite element field solver (PICES), they have simulated a small portion of the MCT.[7,8] This has been used extremely successfully for optimization of device design and processing. They have also created circuit-models of the MCT for in-house use. Other groups have also presented analysis of models dealing with devices similar to MCTs.[9,10]

As a starting point, we have recently developed a simple two-transistor model of the MCT shown in Figure 5. This circuit model is very representative of thyristors and the standard way we think of the MCT device. There are two major problems with this approach: first, the adjustable parameters we have in this model do not always clearly relate to the device physics or experimental measurements, and secondly the physics of the MCT device operation differs somewhat from the representation by the two-transistor model. Eventually we hope to quantify and remedy these discrepancies.

Figure 6 shows a good comparison of the model results and a experimental measurements for one complete turn-on/off cycle. However, this was only after adjustment of the model parameters and does not imply any apriori predictive features of the model. The turn-off structure of the data shows that there is a fast component and a slow component (also seen in other data) not shown by the model. The on-state voltage drop of the model does not change during the pulse as does the experimental voltage drop; this may be a self-heating effect which is not included.

Two-transistor MCT Model
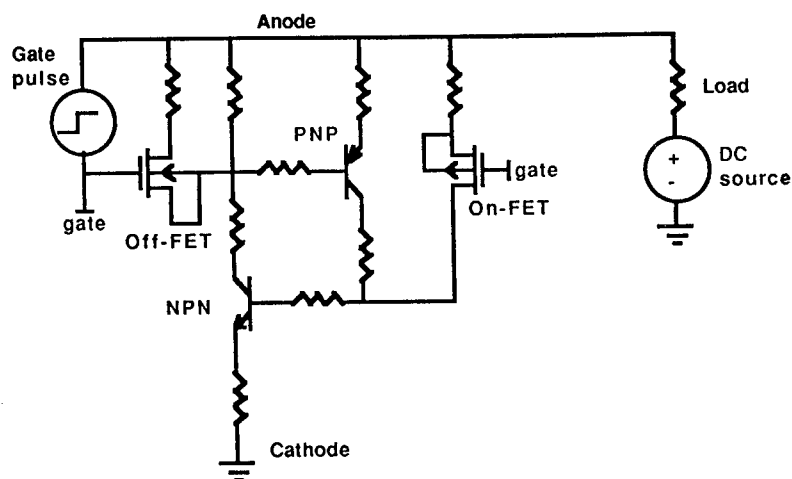
Figure 5. Shown is the circuit diagram for the two-transistor circuit model. The PNP and NPN back-to-back transistors form the thyristor structure. The gate pulse is reference to the anode and when ≥ 2 V turns on the off-FET, when ≤ -2 V on the on-FET.
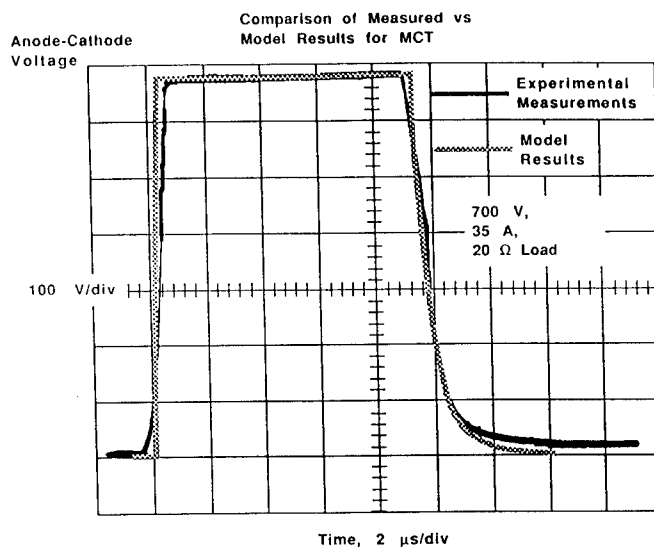
Figure 6. Comparison of measured versus model results for one turn-on/off cycle. Here the device is initially off, then turned on at a time of 2 μs. The measured results were for a 15P90 device; the turn-on time was 0.86 μs, the turn-off time was 1.37 μs.

153

Refinement of the model is proceeding as we test more devices and as the device design stabilizes. This simple model has the advantage of being easy to understand and modify, and runs relatively quickly on a desktop computer. We will seek to incorporate a better representation of the physics of the wide base into the model and add temperature-related effects. In the near future we will use this model to analyze series/parallel arrays of devices for indications of the stresses and possible failure mechanisms. This analysis is important since most high-power applications will require extensive use of arrays.

V. Application Issues

Applications are the driving concern for any device development program. In the high-power arena there are several distinct families of applications each with a different need for the switching components. The military has needs in all these areas as previously discussed in the introduction.

While the specifications will depend on the particular application, we can identify a standard "benchmark" application for each family. The families of applications are broken down to the following: motor controllers (including actuators), inverters, inductive energy storage, modulators, and circuit breakers. Within these groups the switching needs will be similar, although the power levels can differ greatly.

Pulse Width Modulated 100 HP Motor Controller Specifications

Amps On/Off	Voltage volts	Rep Rate	Time On / Off	FVD	Switching Loss	Power Dissip.
400 max 300 ave.	270	1 - 5 kHz	< 10 μs	Very import.	Less important	Very import.

Advances in solid-state motor controllers offer savings in energy and size/weight. Industry and the power utilities are particularly interested in the energy savings since about 20% of the electric power in the United States is consumed by motors. These devices range from a few to one hundred thousand horsepower in applications such as air conditioners, mechanical movement, and pumps. Commercial switching devices (FETs, BJTs) can easily handle low voltage applications, but for high-power applications the MCT holds great promise.[11]

Pulse Width Modulated High-Power Inverter Specifications

Amps On/Off	Voltage volts	Rep Rate	Time On / Off	FVD	Switching Loss	Power Dissip.
1 kA	500	≤ 50 kHz	< 5 μs	Less import.	Very important	Very import.

Inverters are used to transform DC power to AC/DC power at different current and voltages. Often the load requirements for current and voltages are very different from the prime power source. An inverter chops the input DC to a short pulse which is transformed to the desired voltage/current level and filtered to produce the output. The output voltage is regulated by pulse width modulation of the chopping pulses. Inverters/converters range from a couple of watts to megawatts. Generally, the trend is toward high frequency (10s of kilohertz) to reduce the size of the transformer and gain finer control on the voltage regulation. For standard low voltage converters (i.e. 28 V DC from batteries to 120 V AC), FET switches perform very well. For higher input voltages MCTs are attractive due to lower power losses.

EM Gun 10 MJ Inductive Energy Opening Switch Specifications

Amps On/Off	Voltage volts	Rep Rate	Time On / Off	FVD	Switching Loss	Power Dissip.
1MA peak	< 5 kV	10 Hz	100 µs (turn off)	Very import.	Not important	Very import.

Inductors have the potential to store a much larger energy density than capacitors. Especially with the recent advances in high-temperature superconductors, inductive energy storage is very promising. The ideal storage element would store energy without loss, can be rapidly or slowly charged/discharged, and stores a large quantity of energy in a small volume. There is always a demand for greater energy densities; fast discharge capacitors store about 1 J/gm, batteries about 300 J/gm, and inductors about 100 J/gm.

The difficulty here is that batteries can not be rapidly charged or discharged, and that there is an intrinsic energy loss of about 20% which goes to heating. Capacitors can be quickly charged/discharged but have a relatively low energy density. Superconducting inductive energy storage is the ideal energy source with a near infinite storage time, high energy density (almost that of a chemical battery), ability to charge and discharge at any rate.

The limiting element here is a highly efficient opening switch. No energy can be extracted without a switch to open the inductive circuit. For the switch, opening a high current in a large inductor is the worst possible case. Because the current is changing in the inductor the induced voltage $-L\, \partial I / \partial t$ will dramatically rise (i.e., for a 20 µH inductor, 1 MA current, 2 ms opening time, the inductive voltage kick would be 10 kV). Also the inductor will typically be cryogenically cooled - which means the switch must also operate at cryogenic temperatures. Assuming the switch to always be in the current path in the inductor, then the energy storage time is dependent on the losses in the switch. No available switches can meet all the criteria - opening under high currents and voltages, zero forward voltage drop, able to operate at cryogenic temperatures. The MCT is promising for this application, but will still be a limitation.

155

Search Radar 50 kW Microwave Amplifier Modulator Specifications

Amps On/Off	Voltage volts	Rep Rate	Time On./ Off	FVD	Switching Loss	Power Dissip.
1 0	40 kV	1 kHz	≤ 10 µs	Very import.	Less import.	Very import.

Modulators are the driver circuit in microwave generators for applications such as radars, neutral particle beams, free electron lasers, linear accelerators, HPM, etc. The switch applies a high-voltage pulse to the cathode of a microwave tube. This generates an intense burst of electrons which are accelerated and then used as the pump for microwave amplification. The high voltages required, from a few to over a hundred kilovolts, result in use of large series stack of switches. Solid-state switches such as the MCT offer high efficiencies due to low forward voltage drops, and the ability to turn off the cathode pulse in case of a fault condition (saving the tube). Further, the ability of a MCT to turn off allows the radar to dynamically vary the pulse width, repetition rate, and coding. Modulator circuits range from kWs to MWs average power. Solid-state switches are attractive here since the power lost into the switch is about 1/10th of the loss into a standard triode/vacuum tube.

Power Utilities 15 kV, 600A Circuit Breaker Specifications

Amps On/Off	Voltage volts	Rep Rate	Time On / Off	FVD	Switching Loss	Power Dissip.
20 kA max 600 ave	15 kV	100 Hz	≤ 500 µs (off only)	Most import.	Not important	Very import.

Circuit breakers are used to detect and stop excessively high currents, such as would occur during a fault. Traditional circuit breakers rely on mechanical opening switches to interrupt the current. Often this is much too slow (about 10 ms or longer) to prevent damage from occurring. Further, the threshold level is usually quite high to prevent accidental interruption (and thereby system shutdown) and sometimes allows too much energy to the load. For some applications mechanical circuit breakers are too large or heavy; here they must resort to a one-shot fuze which disables the system until a replacement fuze can be inserted. The concept of advanced circuit breakers makes use of high-power solid-state switches to quickly interrupt current before any damage can take place. Further, these solid-state circuit breakers can reset in microseconds enabling systems to continue to function immediately after the fault clears. Since turn off is required, GTOs and MCTs are possible candidates for high-power systems.

All of the above applications requires that the switch have a high efficiency. The energy dissipated in the switch not only takes away from energy to the load but more troublesome it causes substantial heating of the switching elements. This necessitates cooling of the switch since the safe-operating-area of a solid-state device quickly degrades

above a threshold temperature. Cooling systems are very costly due to the size, weight, and complexity they add to the system. Consequently, efficiency in a switching device is a high priority. The MCT, because it is an efficient high-voltage switch and because it can operate at elevated temperatures, is very attractive for use in high-power applications.

VI. Summary

The MCT is a new high-power solid-state switch capable of turning on and off significant currents at high voltage. It has advantages over other devices such as low-forward-voltage drop at high currents, operation under a wide temperature range, ease of control, fast turn on and off. Available solid-state switching components in the high-power arena are quite limited; each has features that may make it best for certain applications. The MCT has the potential of combining the key features of MOS FETs (ease of control and speed) with that of a thyristor structure (high voltage, low forward voltage drop, efficiency) to make a hybrid device that has a wide span of applications. Eventually, we expect this device to displace other high-power turn-off switches for applications above 500 volts.[4,8,11,12]

The experimental measurements we have taken have been used to improve device design and at the same time we gain experience in using MCT devices. We have demonstrated for the first time operation of MCTs at very high repetition rates (50 kHz) and in a series stack holding off over 2,000 volts with five devices rated at 500 V blocking. Our work has shown that MCTs can be safely operated at 80% of their rated DC breakdown voltage when a snubber circuit is used. We have begun modeling MCTs as part of our applications support and have shown that a two-transistor model is usuable as a first approximation to the MCT. Our discussion of applications concludes that efficiency is the most important quality in a switch.

Our work here at the Pulse Power Center has centered on testing, modeling, and technology insertion. Current Army application work includes coordination with the Tank and Automotive Command for developing MCT-based concepts for future tank systems (Block IV) as the prime means of power management, and work with Aberdeen Proving Ground in using the MCT as a reliable fuzing device for warhead testing.

Bibliography

[1] V. Temple, "MOS Controlled Thyristors (MCTs)," *International Electron Devices Meeting Technical Digest,* pp. 282-285, 1984

[2] V. Temple, "The MCT, A New Class of Power Devices," *IEEE Trans. on Electron Devices,* vol ED-33, pp. 1609-1618, 1986

[3] V. Temple, "Power Device Evolution and the MOS Controlled Thyristor," *Power Conversion and Intelligent Motion,* pp. 23-29, November 1987

[4] V. Temple. "MCTs-Thyristor for the Future," *Powertechnics Magazine,* pp. 21-24, November 1989

[5] L. Bovino, S. Schneider, J. Wright, "The MOS Controlled Thyristor (MCT) as an On-Off Capacitor Bank Switch," *Proc. of the 7th Pulse Power Conference,* June 1989

[6] W. Solano, "Low Power MCT Operation," *Auburn University Space Power Institute Report,* October 1989.

[7] V. Temple and F. Holroyd, "Improved Trade-off Between Turn-off Time and Forward Drop," *IEEE Trans. on Electron Devices,* ED-30, pp. 717, 1983

[8] V. Temple, "Advances in MOS Controlled Thryistor Technology and Capability," *Power Conversion and Intelligent Motion,* October 1989

[9] J. Baliga, and H. Chang, "Gate Turn-Off Capability of Depletion-Mode Thyristors," *IEEE Electron Devices Letters,* vol. 10, no. 10, pp. 164-166, 1989

[10] D. Kuo, J. Choi, D. Giandomenico, C. Hu, S. Sapp, K. Sassaman, R. Bregar, "Modeling the Turn-Off Characteristics of the Bipolar-MOS Transistor," *IEEE Electron Devices Letters,* vol. EDL-6, no. 5, pp. 211-214, 1985

[11] R. King, A. Radun, H. Chang, J. Rulison, "Numerical and Experimental Comparison of Power Darlington, IGBT, and MCT Device as a Switch for Adjustable Speed PWM Inverter Drive Applications," *Power Conversion and Intelligent Motion,* October 1989

[12] N. Hingorani, H. Mehta, S. Levy, V. Temple, H. Glascock, "Research Coordination for Power Semiconductor Technology," *Proc. IEEE,* vol. 77, no. 9, pp. 1376-1389, 1989

Laser Microscope using Phase-Conjugate Reconstruction from a Photorefractive Hologram

Philip S. Brody and Charles Garvin
Harry Diamond Laboratories
U.S. Army LABCOM, Adelphi MD 20873

Introduction

In this paper we describe a real-time laser holographic microscope which uses real-time photorefractive holograms in conjunction with real-time digital processing. The new microscope produces images in different types of phase contrast, including a type similar to that produced by the incoherent light method of "differential inter-ference contrast" and a type similar to that produced by conventional "phase con-trast" methods. The microscope produces high-contrast images of even weak phase objects.

The microscope can also be used to image (without digital processing) moving phase components of a phase object. Stationary phase components in the subject do not appear in these images. Additionally, the holographic microscope can be used to obtain distortion-free images of absorptive microscope subjects embedded in a transparent distorting medium.

In this paper we first describe the principles of operation of the laser microscope and then the actual device and its operation. We follow with images of conventional subjects mounted on microscope slides which demonstrate the operation of the microscope.

Principles of Operation

Consider a laser beam transmitted through an object mounted on a microscope slide. The object modulates both the phase and intensity of the plane-wave incident beam, so that the beam on the other side of the slide contains both distortions in the plane-wave fronts and intensity patterns due to optical absorption. Let the beam

then pass into a specifically oriented crystal of barium titanate. A volume phase hologram forms spontaneously in this crystal, the result of photorefractive self-pumping [1,2]. The hologram records the phase and intensity of the transmitted modulated optical field. Once formed, the hologram generates, from the illumination incident upon it, a new beam that is the phase-conjugate reconstruction of the original transmitted beam. This beam is a propagating optical field with surfaces of constant phase (wave fronts) which are identical to those of the original transmitted field. The directions of motion of the fronts are, however, reversed. The new beam back-propagates from the crystal. The beam contains the optical reconstruction of the object which falls superimposed upon the object itself. This is a "real image" reconstruction. It can be viewed, as if it were the object, with an ordinary microscope, using a beam splitter to divert the propagating optical field into the microscope objective (see fig. 1).

When this propagating field passes back though the phase-transforming elements of the object, these elements modulate the propagating optical field in a rather remarkable way. Regions of advanced phase are retarded, and regions of retarded phase are advanced. This happens in such a way that the distortions in the wave fronts which resulted from the initial passage are removed [3,4]. If the original beam was characterized by planar surfaces of constant phase, then the wave fronts in the beam which emerges after the backward passage are again free from distortions with characteristic planar surfaces of constant phase. Such a plane-wave-front optical field contains no intensity patterns due to phase. Intensity patterns due to light absorption remain, with the intensity levels further decreased by the second passage through the object.

Figure 1. Schematic of holographic processing microscope.
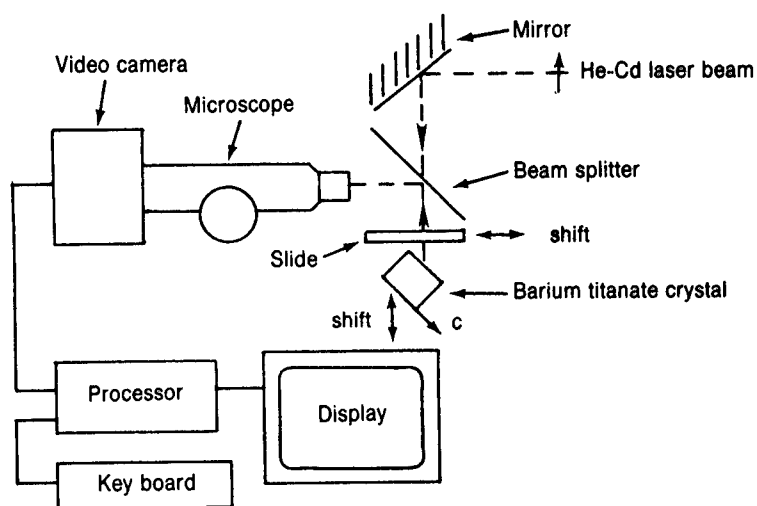
Consider now the effect of transmitting this phase-conjugate optical field back through the object, but only after the object has been displaced some incremental amount. Distortions in the fronts of constant phase will no longer be completely eliminated. Some front distortions will remain, and the degree and character of the distortions will depend on the magnitude and direction of the displacement.

We have studied the effect of displacements both in experiments and analytically. We have found that the resulting intensity variations in the back-propagating beam in the near-field region next to the plane of the object, for small displacements of the object, are phase-contrast patterns of the phase elements in the object [5–7]. Shifts in the plane of the object create patterns which show gradients in introduced phase with respect to the shift direction. Shifts of the object in the direction of the beam also result in phase contrast; these shifts produce patterns which show phase retardation directly, as in "phase-contrast" microscopy.

The patterns created by the back-propagation through a displaced object are the basis of another form of phase-contrast imaging, in which moving phase-retarding elements in an object are visible while stationary elements vanish. Viewing the phase-conjugate optical field again, after it has passed back through the object, we see no intensity patterns in the back-propagating optical field due to the stationary phase elements. The crystal, however, has a finite response time with respect to hologram formation, and the moving elements are therefore constantly displaced with respect to the reconstruction. The reconstructed optical field thus passes through shifted phase-retarding elements; patterns showing these therefore appear in the back-propagating beam.

The phase-contrast patterns that we have been discussing can be imaged with a conventional microscope, and the microscope images can be directly photographed [5,6]. There remains, as in all laser microscopy, a problem with coherent artifacts which appear as a background pattern. The coherent artifacts can, however, be removed by a simple digital processing step. The phase-contrast image appears on the coherent artifact background only as the result of a shift. The background can thus can be removed by subtracting an initial image containing the artifacts from a final image containing the phase-contrast image along with the background [8].

Consider also the direct optical image obtained with the viewing microscope when the object consists of light-absorbing elements immersed in a distorting medium. The object is illuminated by the back-propagating phase-conjugate beam, and as a result distortions in back-propagating wave fronts due to the medium are

161

reduced. This type of distortion removal is not within the capabilities of conventional microscopy.

The Laser Microscope

The configuration for the laser microscope is shown in figure 1. A 10-mW linearly polarized beam (with a wavelength of 441.6 nm) from a helium-cadmium laser passes first through a pellicle beam splitter and then through the object, a specimen mounted on a microscope slide. It then passes into a barium titanate crystal. The crystal is $5 \times 5 \times 7$ mm and oriented so that the beam enters with polarization parallel to the c-axis and at an angle of $45°$ to the positive c-axis. A hologram forms in the crystal by self-pumping within about 20 s. Once formed, however, it can be modified in real time. Moving a different part of the specimen into the beam results in a new hologram, storing information from the second input. The hologram will follow a changing input in a typical specimen, provided the object motion relative to the beam is not too rapid.

The hologram in the crystal now returns the phase conjugate of the transmitted beam which passes back through the specimen to the beam splitter, where it is diverted into the microscope objective. The microscope slide is tilted so that specular reflections from the glass surfaces do not enter the microscope objective. The microscope slide is mounted in a transport that includes two piezoelectric actuators. These are used to shift the slide in its plane (transverse shifts) by a specified distance, in a specified direction. The crystal is mounted on another piezoelectric actuator. The actuator can be used to shift the crystal (and its resident hologram) in the direction of the beam.

We note that the optical field that excites the hologram after the shift is not identical to that which created the hologram, because the optical field has been transmitted through the shifted rather than the initially positioned specimen. The hologram is a selective Bragg reflector and will select (from an input illumination with a sufficiently broad angular spectrum) the angular spectrum components that will generate the phase-conjugate optical field. In practice, the beam when transmitted through a slightly shifted specimen provides these components. If the shift is too great and the spectrum highly modified, the intensity of the returned beam will be significantly reduced.

The objective images a plane in the back-propagating beam onto an intermediate image plane. The ocular reimages this image onto the sensor plane of a video camera. The signal from the camera passes into the image processor, where it is digitized

162

to an eight-bit resolution and stored in a 512 × 512 pixel format. Either the direct video image or a processed image can be displayed on the video monitor.

An image is acquired when a keyboard signal directs the processor to digitize and store an initial microscope image that contains a pattern which includes coherent artifacts. The same keyboard signal also directs the actuators to shift the slide and/or crystal. The shift is too rapid to allow a new hologram to form. The original hologram illuminated by the beam passing through the shifted slide reconstructs the phase conjugate of the beam, phase transformed and intensity modulated by the specimen in its original position. This beam now passes through the shifted slide. The result is a second microscope image that is again digitized and stored, which contains an additional pattern due to the shift. The processor subtracts the first stored image from the second and the result, free of coherent artifacts, is displayed on the video monitor.

The original intent was to produce shifts in the direction of the beam (longitudinal shifts) by shifting the slide in that direction. It was found more convenient, however, to shift the crystal instead. The two are not exactly equivalent operations, because phase features in the beam inherent in the optical system will be imaged in addition to those produced by transmission through the specimen. In practice this did not seem to be an important factor.

Images of moving phase objects which result from the objects' own motion rather than shifting are processed by a continuous subtraction processing function of the processor, in which the video images are digitally subtracted from one another at a rate of 30 frames per second.

The processor can also be used for digitally generated contrast enhancement. Contrast enhancement can be used to improve images created from weak phase objects such as extremely thin tissue sections.

Discussion of Image Formation

Typical specimens are objects which modulate both phase and intensity. Specimens such as thin unstained tissue sections are primarily phase objects. Others, such as stained sections, are phase objects with important details outlined by light-absorbing stains. Patterns due to absorption appear in the optical images both before and after the shift. If the shifts are limited to the transverse or longitudinal resolution of the viewing microscope, then the absorbing regions will show as areas of uniform gray in the subtractively processed image, a result of the subtraction of two

identical optical images. Larger shifts will result in images showing gradients in gray-scale levels.

Phase-modulating structures which refract rays to large angles will not enter the aperture of the phase-conjugating crystal. Such phase-modulating structures of high spatial frequency are beyond the resolution limit of the phase-conjugating process. As a result, patterns showing these appear in the back-propagating field and in the optical image before and after a shift. They then also appear in the processed image, where they look similar to patterns due to optical absorption.

Image contrast and acuity depend on the degree of defocus of the viewing microscope [5]. To produce sharp images, the viewing microscope is focused on the plane of the object, which is also the plane of the reconstruction. It should be noted, however, that if the phase-retarding elements lie only in a single plane, there is no intensity pattern in this plane even after a displacement of the object (in that plane). Intensity patterns due to phase appear only after propagation of the optical field a distance back from the plane. To see a phase-contrast image, one must then focus the viewing microscope on a plane a distance back from the plane of the object. As the plane observed moves back from the plane of the object, image contrast appears. Contrast increases as the plane observed is moved back from the object plane. Acuity, however, is lost, as a result of the defocus. Conventional microscope subjects are however refractively thick, and there is no single plane in which the image vanishes. In practice, the best images are obtained by focusing for maximum acuity.

Images

This section presents images of microscopic objects. The images demonstrate some of the capabilities of the laser microscope. The numerical aperture of the objective in the viewing microscope was 0.3, and the (transverse) resolution of the viewing microscope is thus about 1.5 μm. A 10-power ocular reimaged the image produced by the objective onto the camera sensor plane.

The photographs that follow are from the video monitor and show either the direct optical image as seen by the video camera before a shift, or the digitally processed, coherent-artifact-free, phase-contrast image. Note that a direct optical image showing phase contrast (but not artifact-free) also appears on the video monitor just after the shift, although it rapidly fades from view as the new hologram recording the new optical field replaces the initial hologram.

The first set of images is of exfoliated cheek cells (buccal cells). These are weak phase objects. Figure 2a is a processed phase-contrast image obtained with a diago-

(a)

(b)

(c)
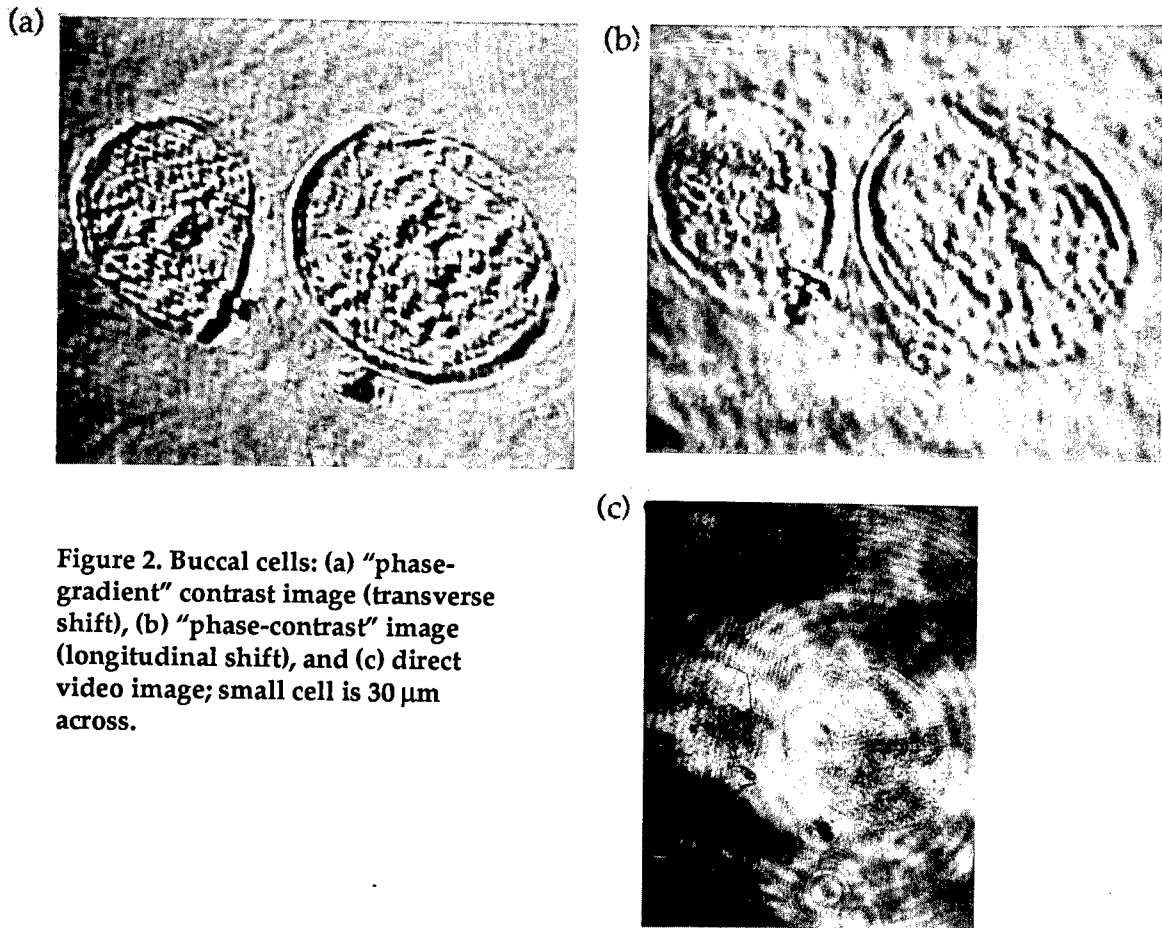
Figure 2. Buccal cells: (a) "phase-gradient" contrast image (transverse shift), (b) "phase-contrast" image (longitudinal shift), and (c) direct video image; small cell is 30 μm across.

nal shift to the right of somewhat less of 2 μm. Figure 2b is a processed phase-contrast image produced with a longitudinal shift of 7 μm. The first shows transverse phase gradients in the direction of the shift; the second shows phase retardation patterns more directly (as in "phase contrast").

Because of the high phase sensitivity of the process, it is difficult to produce either kind of phase-contrast image without some background of phase detail from extraneous sources, such as the fixing medium. The background, however, shows real phase detail and is not an artifact.

Figure 2c shows the unprocessed video image of the buccal cells before a shift. It contains a strong pattern due to coherent artifacts and a weak pattern of phase detail, not resolved by the phase conjugator.

The subject in figure 3a and 3b is a colonial diatom of genus *Fragilaria*. Diatoms are microscopic unicellular algae with cell walls of transparent silica which are

phase objects. The wall consists of two valves (shells) that are more or less flat surfaces held together by a band or girdle. The flat surfaces are not smooth but marked by bumps and ridges which can be used to identify species. Figure 3a is the image taken directly from the video signal before the shift. Figure 3b is the processed phase-contrast image obtained with a 1-μm diagonal shift. The images show the flat ridged surfaces from above. The specimen was stained; the stained regions are the dark central areas in figure 2a and the same areas, now flat gray, in 2b. Ridges which are phase details are clearly seen in the phase-contrast image (2b) and not in the direct bright field video image (2a).

Figure 4a is a processed phase-contrast laser microscope image of diatomaceous earth, a white earthy deposit consisting mainly of fossil diatom shells. The central object is a valve of a diatom of genus *Surirella*. The processed image was obtained using a 2-μm shift in the plane of the slide (down) and a simultaneous 2-μm shift in

(a)   (b)

Figure 3. Colonial diatom, *Fragilaria*, stained specimen: (a) direct video and (b) phase contrast; 12-μm bar.

(a)   (b)

Figure 4. Diatomaceous earth, showing 90-μm *Surirella* valve: (a) phase-contrast image and (b) direct video.
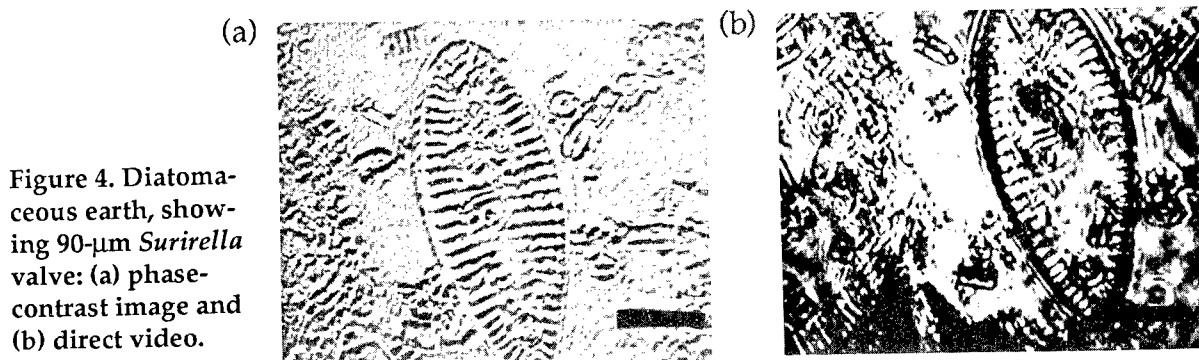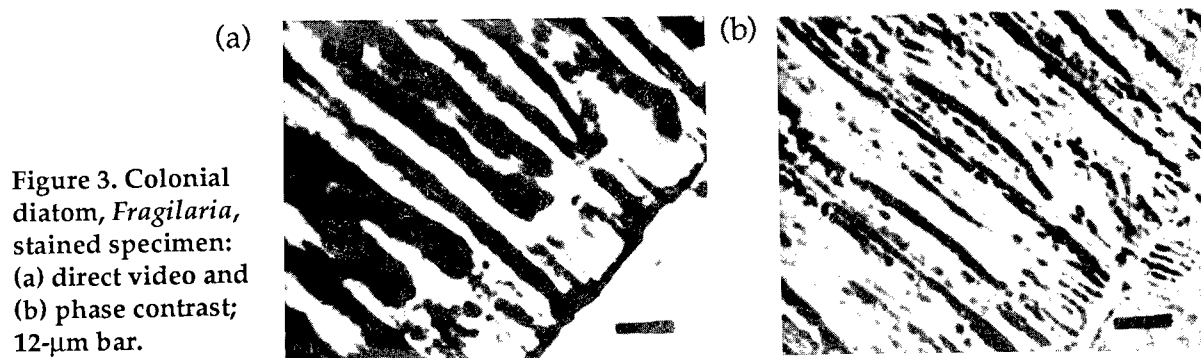
the direction of the beam. Figure 4b is the unprocessed image obtained directly from the video signal, before any shift.

Figure 5a and 5b are processed phase-contrast images of a specimen of the diatom *Eunita faba*. Figure 5a resulted from a diagonally directed transverse shift of 2 μm, and 5b resulted from a transverse horizontal shift of about the same amount. Figure 5c is an image produced with a longitudinal shift (7 μm).

Figure 6 is a phase-contrast laser microscope image of an unstained tissue section from a mouse fetus. The image was obtained with a 1.5-μm transverse shift.

Figure 7 is an image obtained by the same method of one surface of a polished sapphire substrate. A 2-μm diagonal shift was used. One can focus through the substrate and obtain a similar image of the other surface.

(a)
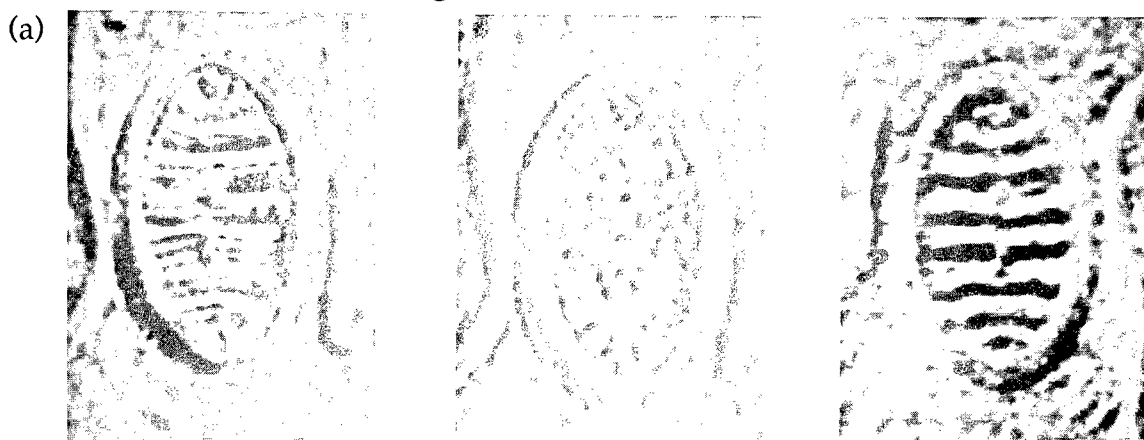
Figure 5. *Eunita faba*. Image produced with (a) a diagonal shift, (b) a horizontal shift, and (c) a longitudinal shift. Long dimension of diatom is 34 μm.

Figure 6. Unstained fetal mouse tissue section, 6 μm thick. Field of view, 200 × 260 μm.
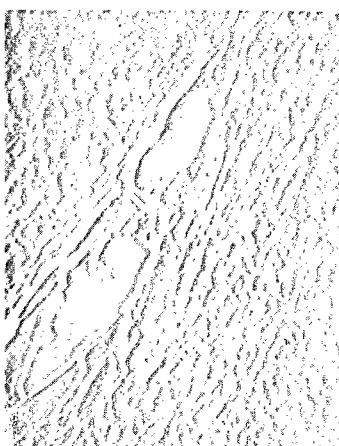
Figure 7. Image of sapphire substrate surface. Bar is 20 μm.

167
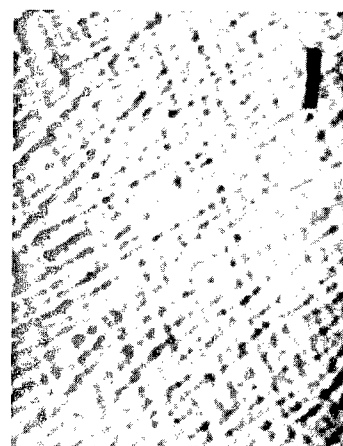
We can demonstrate the effect of the phase-conjugate illumination with respect to removing distortions of light-absorbing elements from a surrounding aberrating phase medium by first examining brine shrimp eggs (which strongly absorb the laser illumination) buried in stopcock grease with an ordinary incoherent light microscope. The bright field image is strongly distorted because of the refractive effect of the grease (Figure 8a). Illumination with the phase-conjugate beam removes the distortion created by the aberrating grease medium (Figure 8b).

Figure 9 is taken from a videotape frame. It is a real-time processed phase-contrast image resulting from the motion of motile single-cell algae in a drop of pond water. The processing is continuous subtraction at 30 frames per second. The processing removes not only coherent artifacts but all stationary background patterns. The sensitivity to phase is great enough so that the water motion due to the agitation of the medium by the swimming algae is imaged.

Figure 8. Brine shrimp egg embedded in grease: (a) bright field incoherent light image and (b) distortion-free image using phase-conjugate illumination. Egg is 200 μm in diameter.
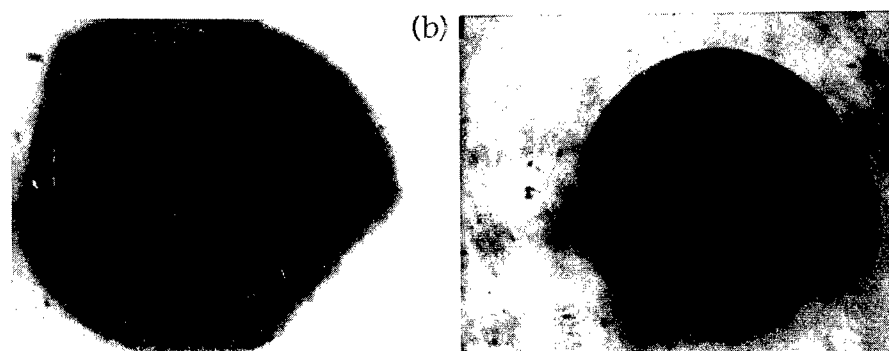
Figure 9. Motile algae in a drop of pond water.
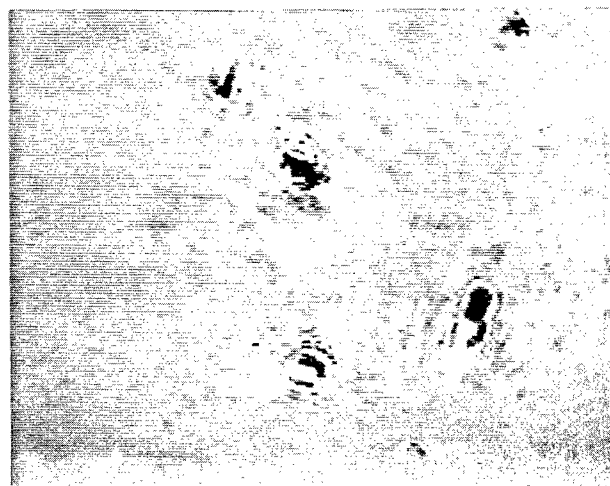
Concluding Discussion

Considerable progress has been made in the development of a laser microscope which uses a phase-conjugate reconstruction from a real-time photorefractive hologram as its illumination source. The device has a number of demonstrated modes of operation, producing a variety of image types. The microscope has some unusual capabilities. On the basis of these capabilities, further development seems to be technically warranted. Developments could further increase effective resolution and contrast, ease of operation, and versatility, resulting in a generally useful laboratory instrument.

References

1. J. Feinberg, "Self-Pumped Continuous Wave Phase Conjugator using Internal Reflections," Opt. Lett. **7**, 486 (1982).

2. P. S. Brody and J. R. Goff, "Grating Evolution and Form in a Single Crystal Self-Pumped Barium Titanate Phase Conjugator," Proc. SPIE **738**, 51 (1987).

3. H. Kogelnik, "Holographic Image Projection Through Inhomogeneous Media," Bell System Tech. J. **44**, 2451 (1965).

4. E. N. Leith and J. Upatnieks, "Holographic Imagery Through Diffusing Media," J. Opt. Soc. Am. **56**, 523 (1966).

5. P. S. Brody and R. P Leavitt, "Dynamic Holographic Method of Imaging Phase Objects," Appl. Opt. **26**, 913 (1987).

6. P. S. Brody and John D. Bruno, "Intensity Imaging of a Phase Object and Auto-correlation using a Barium Titanate Optical Phase Conjugator," Proc. SPIE **754** (1987).

7. J. D. Bruno, B. D. Wong, and P. S. Brody, "Diffraction of Light by a Phase Object," Harry Diamond Laboratories unpublished report (1987).

8. P. S. Brody and C. Garvin, "Microscopy Using Optical Phase Conjugation in a Hybrid Analog/Digital System," Proc. SPIE **901**, 119 (1988).

The Effect of Crystallographic Orientation on the
Performance of Single Crystal Tungsten Sub-Scale Penetrators

William J. Bruchey, Dr.
Edward J. Horwath, Mr.
Priscilla W. Kingman, Dr.
Ballistic Research Laboratrory
Aberdeen Proving Ground, MD 21005-5066

1. INTRODUCTION

Various high density materials have found usage in long rod kinetic energy (KE) projectiles. These materials usually include tungsten-nickel-iron composites with 90% or more tungsten, and uranium alloys such as U-3/4 Ti. Attempts to improve kinetic energy penetrator performance have centered on improving the tensile or compressive strengths of these tungsten or uranium materials. These performance improvement attempts have been less than fully successful partially due to a lack of understanding of the basic high deformation rate failure mechanisms of the materials. A starting point to understanding the behavior of these alloy or composite systems is to first fully understand the individual component behavior.

Consequently, single crystals of tungsten were selected for investigation as penetrator materials. A model scale program was initiated to determine the mechanisms influencing penetration of tungsten single crystals of various orientations ([111], [110], [100]). An expected outcome of this work was to determine if crystallographic orientation and subsequent materials deformation systems associated with a unique crystal orientation had an influence on the penetration process.

2. EXPERIMENTAL PROCEDURE

Single crystals of tungsten in three orientations [111], [110], and [100] were obtained from Atomergic Chemetals

Corp., Farmingdale, NY. The crystals were grown using the Czochralski (CZ) technique after a preliminary two pass zone refinement of 99.99 purity starting material. This growing technique allowed production of crystals with no more then 1 degree deviation from the above orientations, and with dislocation densities less than $1 \times 10^7 cm^{-2}$. The crystals were lightly centerless ground and then chemically etched (weak HCL solution) to produce strain free crystal surfaces. The crystals were obtained in a L/D = 15 sub-scale penetrator geometry with 102.5 mm length and 6.90 mm Dia. with a weight of approximately 74 grams. One crystal of each orientation was back reflection Laue x-rayed to determine the perfection of the crystals. In all cases the perfection was good, with minimal pole distortion, and the orientation was considered to be within the tolerances claimed by the manufacturer. Additionally, compressive tests were completed on .050 inch diameter crystals produced by the manufacturer using the same CZ growing technique. Compression data are presented in table 1.

TABLE 1. Single Crystal Compression Data

Orientation	Offset Yield Strength (ksi)		
	.2%	1.0%	5.0%
100	52	102	193
110	95	102	132
111	112	151	266

The compressive data of Table 1 are fairly consistent with other published mechanical properties information on tungsten single crystals [1,2,3,4,5]. The [110] orientation had limited work hardening capability, and lower room temperature failure loads.

The following four (4) "state of the art" subscale penetrators were also obtained for performance comparison: 1) Geometry equivalent 90-7-3 tungsten nickel iron composite rods with 102.5 mm length and 6.90 mm Dia. (65 gram weight); 2) Mass equivalent and L/D = 15 equivalent, 74 gram 93% tungsten alloy; 3) Mass equivalent and L/D = 15 equivalent, 74 gram uranium 3/4 titanium rod processed to standard army penetrator material specification; 4) Mass equivalent and diameter equivalent, 74

gram 93% tungsten alloy rod.

All single crystal or alloy rods were launched from a 26 mm smooth bore Mann barrel. A push launched, saboted geometry was used in all test cases. Two target geometries were selected for performance evaluation of the various penetrator rods: 1) Finite 76mm RHA block of hardness Brinell 250 – 275; 2) Semi-infinite RHA block of hardness Brinell 250 – 275. Striking velocities of 1400 meters/sec and 1500 meters/sec, respectively, were used for the finite and semi infinite test firings. If possible, at least two tests were conducted for each penetrator material and geometry combination for each of the target types described above.

Finite tests were instrumented to provide residual x-rays of rod length and velocity for performance discrimination. Pertinent semi-infinite performance discriminator was penetration depth. All penetration tests were flash x-ray instrumented to provide striking velocity and pitch/yaw. After penetration test, samples were removed for metallographic observation with light and scanning electron microscopes, as well as x-ray diffraction equipment.

3. BALLISTIC RESULTS

The push launch of these single crystals was in general very successful. Some launch breakage problems were encountered, especially for the [110] single crystal orientation. Additionally, one of the [100] tungsten single crystal semi-infinite tests rods failed on launch, and the limited number of single crystals did not allow repeat of this penetration test. Failure of the [110] orientation crystals on launch is related to the lack of crystal ductility in this orientation, as well as a [110] cleavage plane oriented perpendicular to the stress axis. What appeared to be overload cleavage failures were observed at 90^0 degrees to the penetrator axis in the striking flash radiographs. Tensile test at low temperature (77^0K) by Argon and Maloof [3], indicate that the elongation to failure of the [110] orientation is considerably less than the other two orientations. This same brittle behavior exhibited at low temperature in the [110] orientation can be expected at high strain rates. It is believed the [100] launch failure was related to an undetected crystal growing defect, and is not characteristic of the behavior of this crystal orientation.

Test data for the ballistic penetration tests (finite and semi-infinite) are presented in Table 2. The ballistic data presented in Table 2, for semi-infinite testing, show both raw test data and penetration numbers corrected for velocity discrepancies (based on approximately 1 mm performance difference for every 10 meters/sec velocity correction [6]). The commonly used ratio of semi-infinite penetration to penetrator length (P/L) is also presented in this table. The ballistic results of this testing, presented in Table 2, indicate that the performance of these tungsten single crystals is: 1) strongly orientation dependent; 2) approximately equivalent to uranium 3/4 titanium. The [100] orientation tungsten single crystal rods had approximately equivalent semi-infinite and finite ballistic test performance to uranium 3/4 titanium. The [111] tungsten single crystal orientation had poorer performance than the U-3/4Ti or [100] orientation, but still had P/L performance better than 93% tungsten alloys.

Tracings of the semi-infinite penetration craters, and finite residual penetrators are presented in Figure 1 and Figure 2, respectively. The [100] penetration cavity walls, Figure 1, are much smoother than either of the other tungsten orientations. The [100] orientation walls are also much smoother than the tungsten alloy or uranium 3/4 titanium walls. The residual penetrator masses for [100] oriented tungsten single crystals and uranium 3/4 titanium appear to be nearly equivalent, with the [100] finite residual broken into three discrete pieces.

The next section of this paper presents some preliminary materials investigative work concerning these interesting ballistic results.

4. PENETRATION BEHAVIOR

After completion of ballistic test, the semi-infinite samples which contained residual penetrator material were EDM machined at the penetration cavity centerline. These samples were rough ground with silicon carbide paper to 600 grit, and then final polished with 6 then 1 micron diamond powders. Photomicrographs of the penetrator cavities and residual penetrators, are presented in Figures 3, 4, and 5. As evidenced in the photographs, sample relief from polishing provided adequate microstructural detail, so no etchants were used. Additional higher magnification photomicrographs of each penetration interface are presented in Figures 6, 8, and 10.

BRUCHEY; HORWATH; KINGMAN

Table 2. Penetration Test Summary

Penetrator/Mass/Length material/grams/mm	Target mm RHA	Vs m/s	Pitch	Yaw	Vr m/s	Dp mm	Dp'	P/L
90-7-3W/64.2/102.5	76.2	1440	0	-.25	684			
90-7-3W/65.1/102.5	76.2	1364	-.25	+.25	125			
93GTE/74.9/106.6	76.2	1420	+1.25	+.75	843			
93GTE/74.4/113.1	76.2	1362	+.50	+.50	774			
93GTE/75.5/106.6	152	1493	+.50	--		89.7	89.7	.84
93GTE/74.5/112.9	152	1493	0	-.25		93.5	93.5	.83
U.75Ti/73.4/104.1	76.2	1439	+.50	-.75	1125			
U.75Ti/74.0/104.1	76.2	1356	-.25	-.25	1097			
U.75Ti/74.0/104.4	152	1468	-.50	+1.0		97.0	99.0	.95
U.75Ti/74.0/104.4	152	1483	-1.25	+.25		98.4	99.4	.95
111W/74.1/102.5	76.2	1387	+.75	+.25	855	broke on launch		
110W/73.9/102.5	76.2	1380	0	+.25	323	broke on launch		
110W/73.9/102.5	76.2	1386	+1.5	+.50	PP			
100W/73.7/102.5	76.2	1387	+.50	-.75	1050			
111W/74.2/102.5	152	1485	-.75	-.75		89.9	89.9	.88
111W/73.7/102.5	152	1410	-.25	+.25		82.3	90.3	.88
110W/73.7/102.5	152	1453	+.50	+.50		79.0	83.0	.81
110W/73.6/102.5	152	1511	-.25	+.25		90.0	88.0	.86
100W/74.2/102.5	152	1485	-.25	+.50		99.8	99.8	.97
100W/73.4/102.5	152	1491	0	+.25		91	broke on launch	

(U)

[111]W   [110]W   [100]W   93WA L/D=15   U-3/4Ti
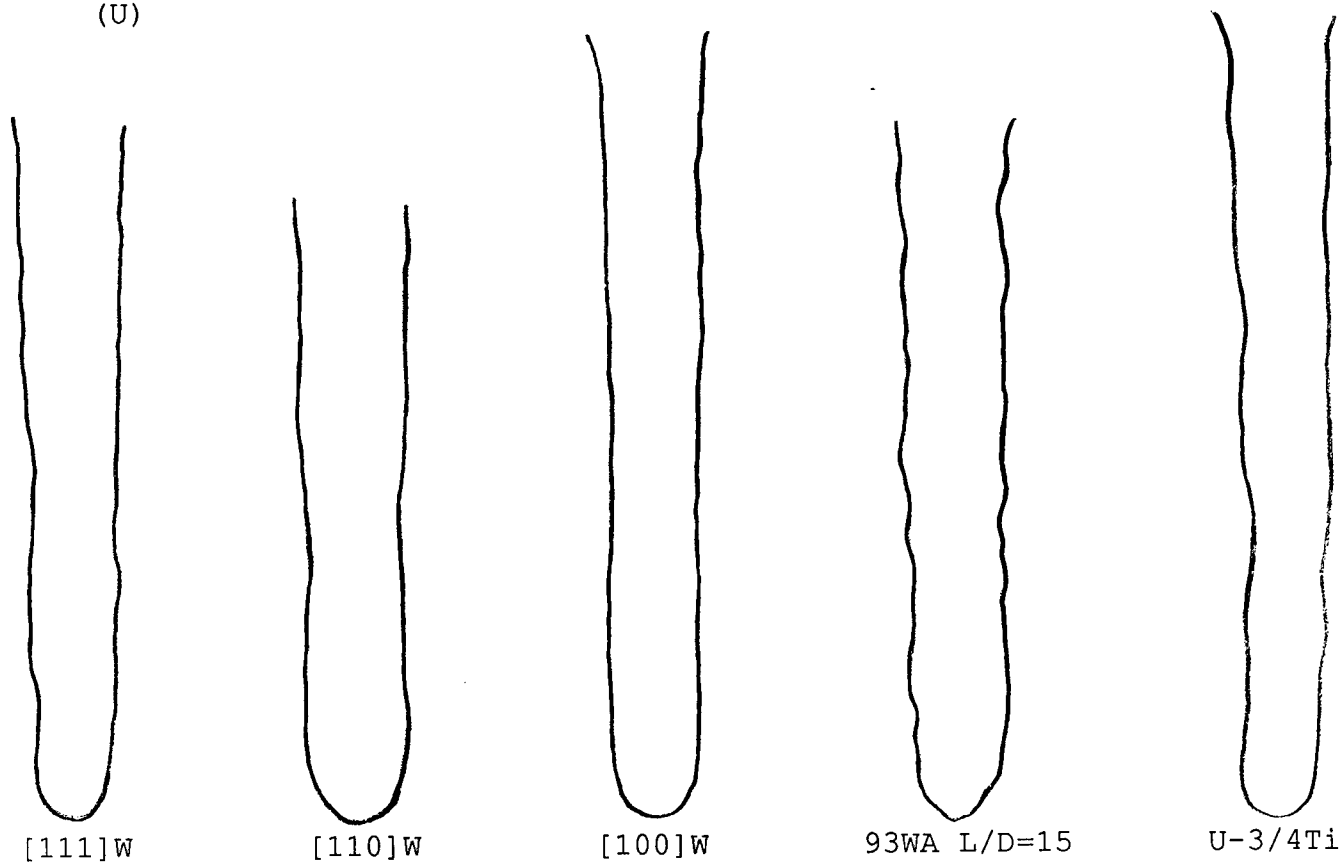
Figure 1. (U) Semi-infinite Crater Profiles

(U)

(U)

[111]W   [110]W   [100]W   93WA L/D=15   U-3/4Ti
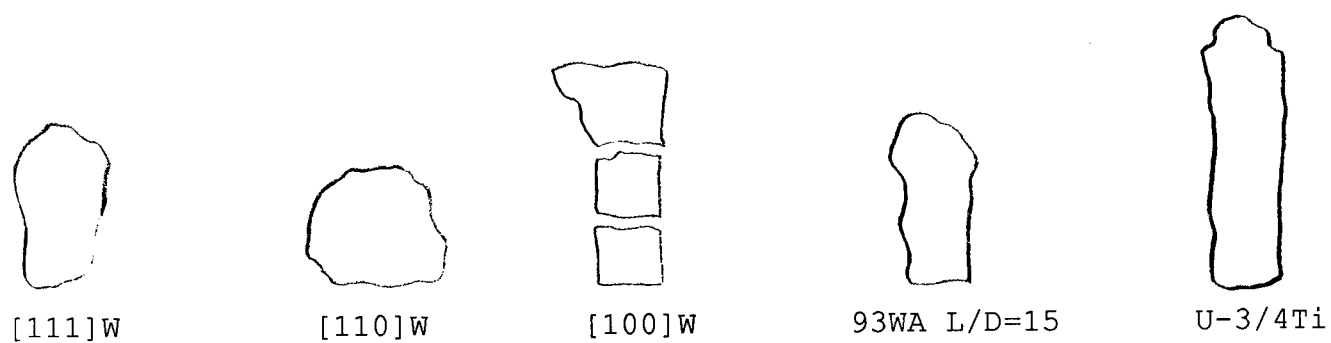
(U)

Figure 2. (U) Finite Residual Penetrator Tracings
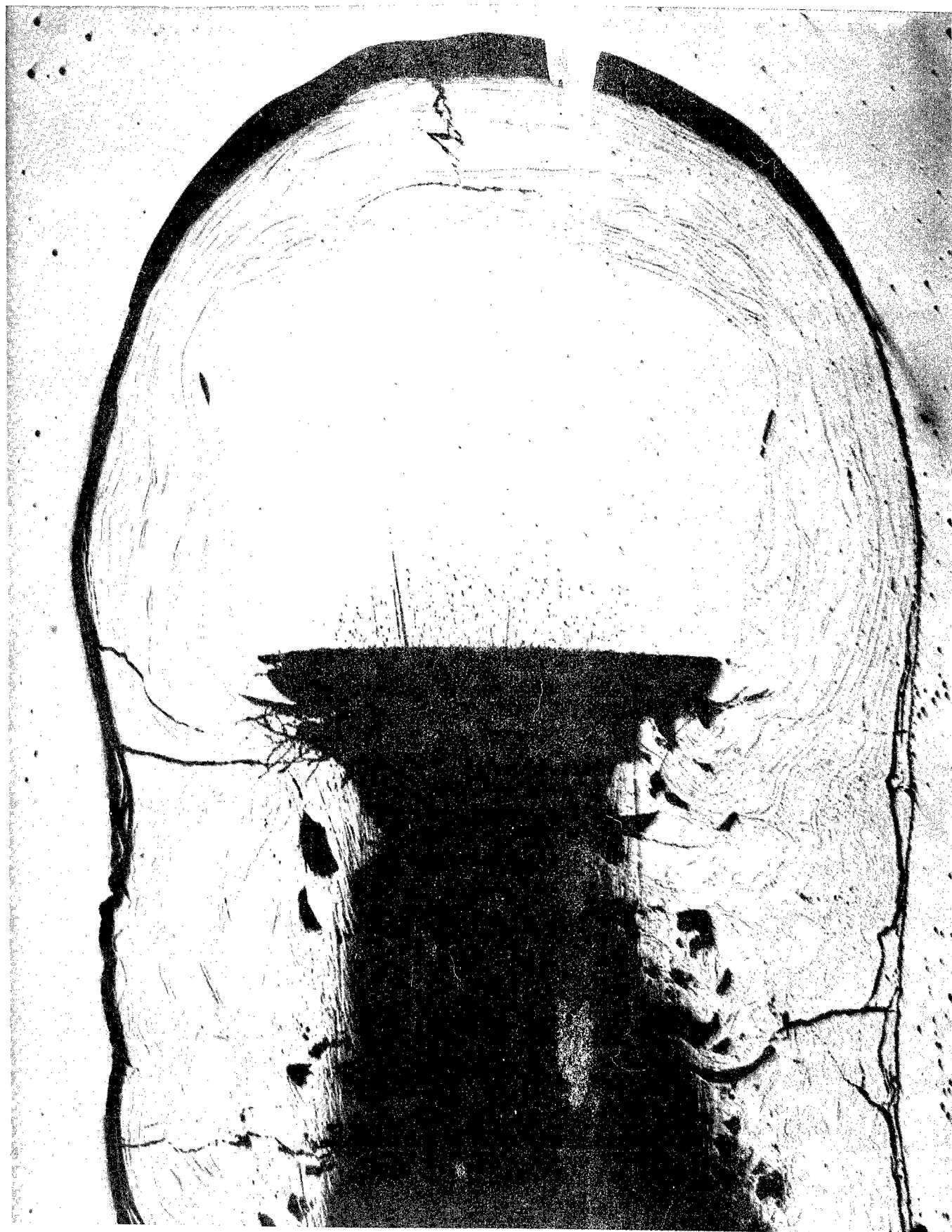
Figure 3. (U) [111] Semi-infinite Residual Penetrator Section
10 X Magnification   As Polished

177

Figure 4. (U) [110] Semi-infinite Residual Penetrator Section
10 X Magnification As Polished

Figure 5. (U) [100] Semi-infinite   Residual Penetrator Section
10 X Magnification   As Polished

179
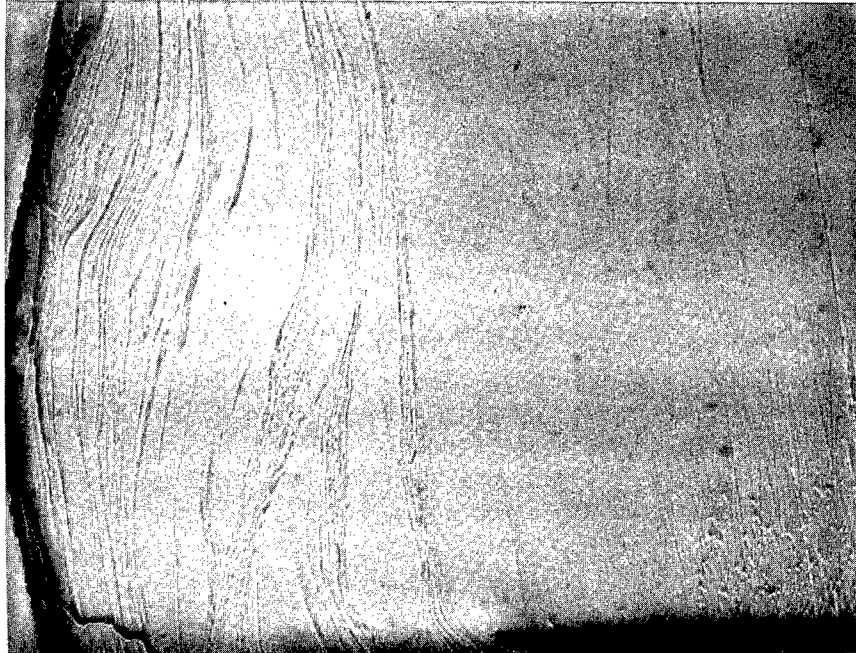
Figure 6. (U) [111] **Penetrator** Residual
25 X Magnification   unetched
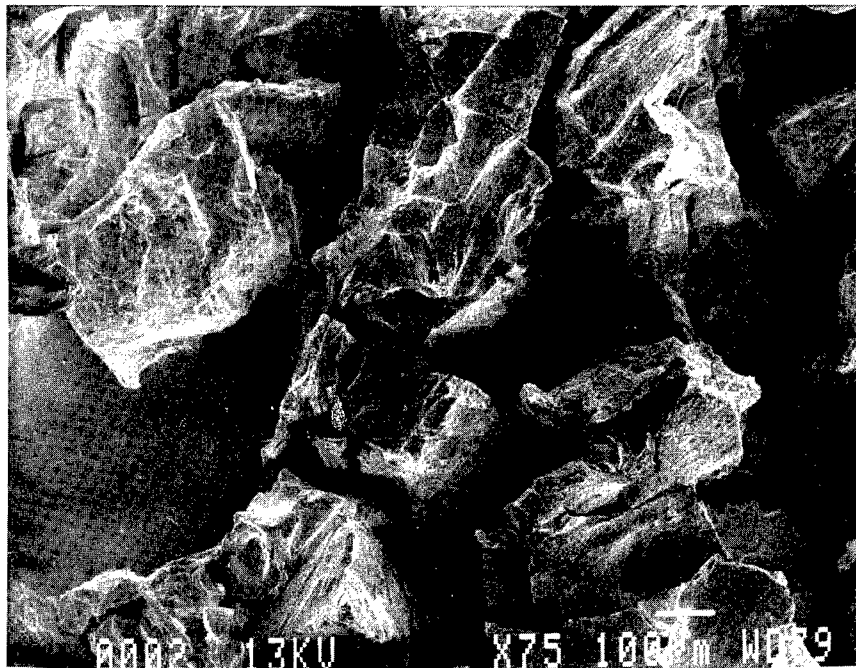
Figure 7. (U) [111] Penetrator Back Extrusion Surface
75 X Magnification   unetched
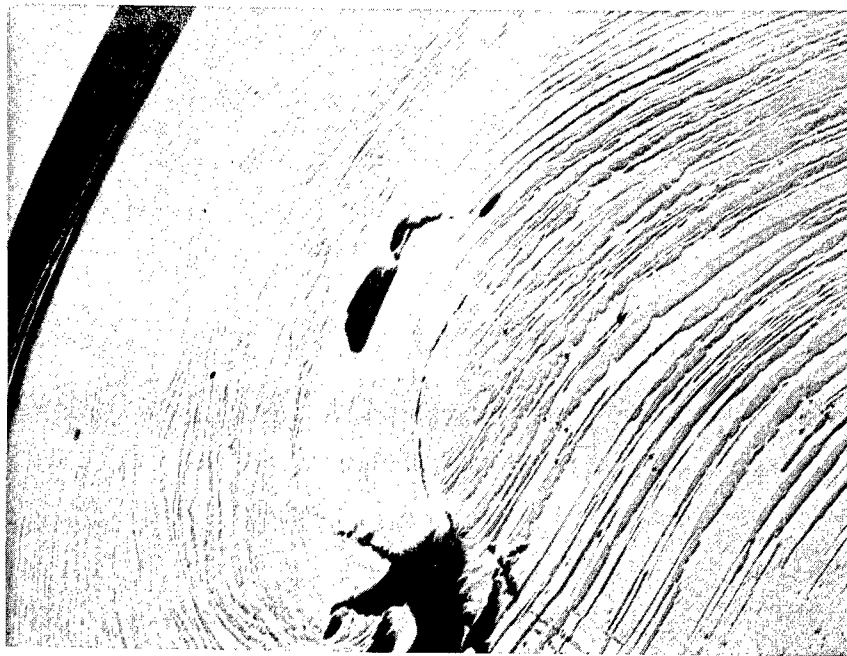
Figure 8. (U) [110] Penetrator Residual
25 X Magnification   unetched
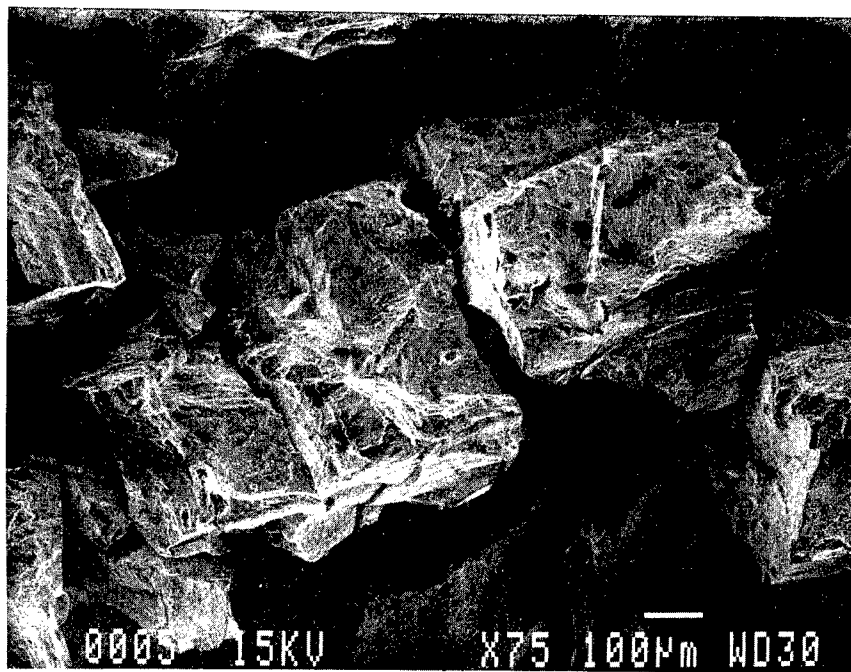
Figure 9. (U) [110] Penetrator Back Extrusion Surface
75 X Magnification   unetched

Figure 10. (U) [100] Penetrator Residual
25 X Magnification   unetched
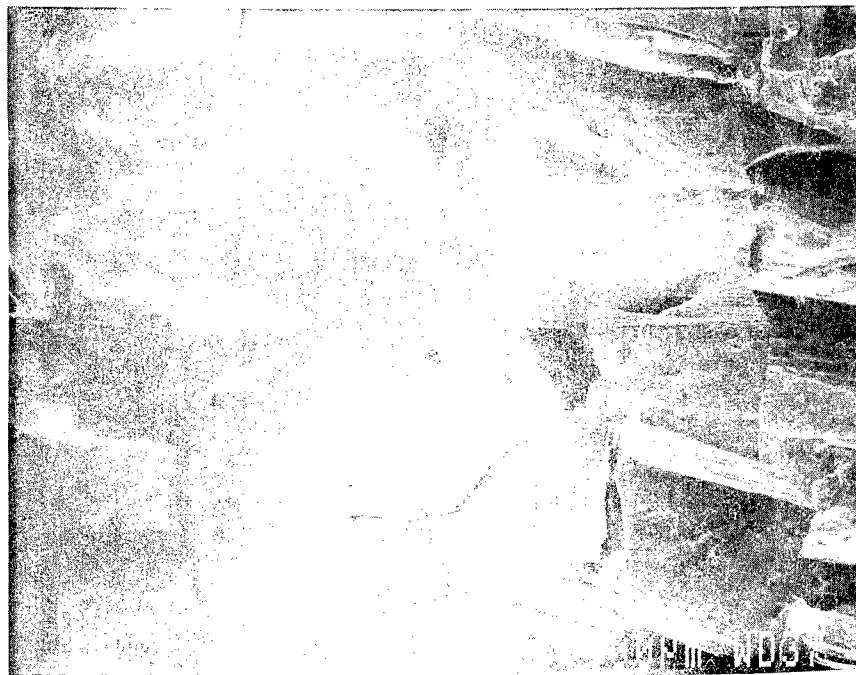
Figure 11. (U) [100] Penetrator Back Extrusion Surface
75 X Magnification   unetched

The photomicrographs display the unique nature of each residual penetrator microstructure and help to elucidate the differences in penetration mechanisms associated with a particular crystallographic orientation.

[111] Orientation - This orientation single crystal had intermediate penetration performance. The microstructure of the residual penetrator is characterized by extensively slipped and worked material surrounding an essentially undisturbed residual penetrator slug. Several slip localizations are evident in the residual penetrator. The penetration/back extrusion process had the following stages: 1) initial dislocation generation and micro-deformation of penetrator in the highly compressed area near the penetration interface; 2) development of a few shear localizations at angle of 12 -15 degrees characteristic for this crystal orientation and slip system; 3) additional shear localization and extensive plastic deformation of material as flow back past remaining penetrator occurs. Generation of relatively few major shear localizations caused a more blocky flow and a non smooth penetration cavity, see Figure 2, 3, and 6. Large regions of nearly undisturbed material are evident within the back extruded sample regions. An SEM photomicrograph of the interior surface of this back extruded material shows the random cleavage fracture structure associated with this single crystal orientation.

[110] Orientation - This orientation crystal had poorest penetration performance. The microstructure is again characterized by extensively slipped and worked material. This orientation appears to generate a large number of slip localizations early in the penetration process. The low compressive strength, Table 1, and lack of work hardening may contribute to the easy slip of the material. The penetration/back extrusion process had the folllowing stages: 1)micro-deformation with dislocation generation and movement; 2) rapid shear localization at angle of 12- 15 degrees, related to inability of material to work harden by dislocation generation in this orientation; 3) movement of material away from penetration interface along the many shear localizations generated. Large number of localizations allowed fairly easy material flow, resulting in a fairly smooth penetratin cavity, Figure 2. The back extruded material structure, shown in the SEM photomicrograph of Figure 9, had a random cleavage type structure.

[100] Orientation – This orientation had the best penetration performance. The material flow was macroscopically very homogeneous late in the back extrusion process. This crystal orientation allowed easy generation of screw dislocations and or micro slip and allowed microscopic deformation to proceed to large strains before large scale slip/shear localization occurred. The penetration back extrusion process had the following stages: 1) large scale microscopic deformation to large strains; 2) final unique shear localization at favorable angle (70 degrees) for easy slip of material and completion of back extrusion. A photomicrograph of the final localization area for this [100] orientation is presented as Figure 10. This orientation produced almost perfectly smooth penetration cavities, Figure 2. The back extruded material had a blocky and very crystallographic appearance, see Figure 11 characteristic of cleavage on the favored {100} or {110} planes [7].

5. CONCLUSIONS

The penetration process of these tungsten single crystal rods is affected by a combination of factors. The compressive mechanical strength of the material contributes to the overall penetrator performance. However, the material strength alone does not predict the penetration performance of these materials. As suggested by Bruchey and Glass, and others,[8] ,the dissipation of initial rod striking energy ($1/2MV^2$) is partitioned as: a) projectile deformation energy; b) crater formation energy; c) strain energy in the target. The evidence from this investigation would suggest that the single crystal orientation influences the back extrusion deformation process and subsequently, the energy loss to deformation of the tungsten single crystal penetrator material. The amount of homogeneous deformation before slip localization and the final slip localization angle were important factors affecting the work needed for this back extrusion of penetrator material.

The deformation process, including homogeneous deformation and slip localization of the penetrator material back extrusions, are then a function of: 1) the penetrator axis orientation relative to the available slip systems in the BCC tungsten structure; 2) The ability of a crystal orientation to generate screw dislocations and then move them through the crystal lattice. The superior performance of the [100] tungsten single crystal orientation, with P/L comparable to uranium 3/4

titanium alloys, occurred because of the very large amount of homogeneous micro-deformation which preceeded eventual slip localization at a favorable angle.

ACKNOWLEDGEMENTS

REFERENCES

1. Rose, R., Ferriss, D., Wulf, J, Yielding and Plastic Flow in Single Crystals of Tungsten, Transactions of the Metallurgical Society of AIME, Vol 224, Oct 1962, p 981.

2. Garlick, R., Probst, H., Investigation of Room Temperature Slip In Zone-Melted Tungsten Single Crystals, Transactions of the Metallurgical Society of AIME, Vol 230, Aug 1964, p1120.

3. Argon, A., Maloof, S., Plastic Deformation of Tungsten Single Crystals at Low Temperatures, Acta Metallurgica, Vol 14, Nov 1966, p1449.

4. Beardmore, P., Hill, D., Deformation and Fracture of Tungsten Single Crystals, J. Less-Common Metals, Vol 9, 1965, p 168.

5. Oku, T., Galligan, J.M., Plastic Deformation of Zone-Refined Tungsten between 4.2 and 300 ^{0}K, Japan Atomic Energy Research Institute, Memo #4124, Aug 1970.

6. Magness, L., Deformation Behavior and Its Relationship to the Penetration Performance of High-Density KE Penetrator Materials, To be published as 1990 Army Science Conference Paper.

7. Tyson, W., Ayres, R., Stein, D., Anisotropy of Cleavage in BCC Transition Metals, Acta Metallurgica, Vol 21, May 1973, p621.

8. Bruchey, W., Glass, J., An Orthogonal Grid Technique for Studying Internal Deformation and Energy Absorption in Target Penetration, Ballistic Research Laboratory Memorandum Report # BRL-MR-3415, Dec 1984.

BRUTTIG, BICKELL, DOHERTY et al.

TITLE: The Use of a Clinically Relevant Uncontrolled Hemorrhage Model to Study Massive Internal Bleeding and to Develop Appropriate Courses of Clinical Response to That Hemorrhage (U).

*MAJ Stephen P. Bruttig, Dr. William H. Bickell, Ms. Tammy J. Doherty, Dr. John D. O'Benar, Mr. Gregory A. Millnamow, Dr. Charles E. Wade, Dr. Marcos Intaglietta, Dr. Per Borgstrom and Dr. Karl -E. Arfors.

Letterman Army Institute of Research, Presidio of San Francisco, CA, 94129-6800, Saint Francis Hospital, Tulsa, OK, 74136-R012, University of California San Diego, La, Jolla, CA, 92093-1992, and Pharmacia Experimental Medicine, La Jolla, CA, 92037

INTRODUCTION

Massive hemorrhage and resulting hypotensive shock have always been a prime concern of emergency medicine as well as a subject of basic scientific interest. Consequently, experimental studies have attempted to provide a meaningful understanding of the response to hemorrhage, such that appropriate clinical therapies could be developed and implemented.[1] Paramount to any scientific investigation of a phenomenon such as hemorrhage is the design and principle of control over the experiment;[2] i.e., the conditions of the experiment are finely regulated, and the investigator studies the response (and its variability) to the controlled set of conditions. For example, the "standard of care" today for fluid volume resuscitation of massive hemorrhage[3-6] evolved from an understanding of the response of a variety of animal models to various forms of controlled hemorrhage,[1,2] as well as from clinical observations in man[7].

While all of the hemorrhage treated clinically represents uncontrolled bleeding (at least initially), it is ironic that little effort in developing proper treatment of hemorrhage has come from the study of uncontrolled hemorrhage. Rather, the preponderance of effort has come from the careful study of

BRUTTIG, BICKELL, DOHERTY et al.

controlled blood loss from catheters inserted into either an artery or a vein. Control over the hemorrhage conditions usually takes two forms; either bleeding to a fixed pressure for a measured period of time, or blood loss induced by the removal of a fixed volume of blood.[1] Either method results in a decrease in blood volume and a significant reduction in blood pressure. The accompanying decrease in blood pressure means that blood flow to many tissue beds falls (i.e., a decrease in tissue perfusion), tissue oxygenation falls and metabolic waste products accumulate. Consequently, a reasonable treatment might be to increase blood volume, which should raise blood pressure. The increase in pressure would effect greater blood flow through tissue beds, thereby facilitating the removal of metabolic wastes, and increasing oxygen delivery to the tissue.

Experimental and clinical experience indicates that certain responses to hemorrhage (cf. above) are reasonably accurate even in man. Consequently, various fluid volume resuscitation schemes have been offered as appropriate treatment therapies for "hemorrhagic shock".[1,3-6] As a result, mean arterial pressure (average blood pressure) was quickly accepted as the endpoint for assessing adequate resuscitation. Initial efforts to match the volume of fluid which would be administered to the estimated amount of blood which had been lost were ineffective, because large volumes of fluid leaked out of the capillaries into the surrounding tissues.[7] As a result, when fluid resuscitation is begun, usually with an isotonic crystalloid solution such as lactated Ringer's solution, blood pressure is monitored. The general idea is that if blood pressure fails to rise to normal levels, more fluid must be given (a general rule is that three times the estimated blood loss must be replaced by resuscitation fluid to compensate for the aforementioned leakage). However, in the clinical situation, post-resuscitation arterial pressure may never reach normal levels (i.e., 90-100 mmHg). As a result, there is a tendency to over resuscitate patients (i.e., give too much fluid) when blood pressure is the pivotal variable.

As early as 1918,[8] and probably long before that, occasional clinicians or basic scientists began to question the validity of the experimental models for hemorrhage and shock as inappropriate to the situation observed in the hospital. Undaunted by these criticisms, a plethora of additional hemorrhage models have been developed during the

BRUTTIG, BICKELL, DOHERTY et al.

intervening years.[1,2,3,9,10] Eventually, the study of hemorrhage attracted the attention of microvascular physiologists, who can observe local blood flow in transilluminated tissues with the aid of special (intravital) microscopes. These investigators have studied hemorrhage both as a subject of scientific interest in its own right[11,12] and as a tool to study the effects of low blood pressure in various microvessels.[13] These studies, while increasing our general understanding of the response to hemorrhage, are themselves a function of the model used to initiate or maintain hemorrhage. As a result, while the findings are interesting,[2] they may only apply to the type of hemorrhage under investigation.

Proposed solutions to the perceived problems. To resolve controversies between the historic literature and current clinical as well as experimental observations, we established a multi-disciplinary program to characterize uncontrolled hemorrhage responses and to determine the efficacy of "standard of care" versus "novel approach" treatments. Our concerns were those regarding the injuries of the combat casualty. These injuries often include disruption of large blood vessels in either the abdomen or the chest (truncal vessels), followed by uncontrolled hemorrhage. Therefore, we sought to develop a model of hemorrhage relevant to our clinical (medical) concerns, and to study both the whole-body (systemic) and microvascular responses to this type of hemorrhage. In addition, we sought to investigate the conditions under which fluid volume resuscitation is usually administered, to define resuscitative therapies for uncontrolled hemorrhage and hopefully to improve the recovery of the combat casualty. To minimize our use of animals and to maximize our total research effort in this endeavor, we sought to develop a simple, but effective computer model for speculating on proposed treatments or estimating "unmeasurable" cardiovascular parameters. Finally, when it became obvious that immediate, aggressive fluid volume resuscitation following massive uncontrolled truncal hemorrhage was often associated with a fatal outcome,[3,9,14] we sought to decrease mortality by delaying fluid volume resuscitative therapy.

MATERIALS AND METHODS

Experimental Animals. Initial development of the model of uncontrolled aortic hemorrhage utilized immature Yorkshire pigs (35-45 kg), surgically instrumented under halothane

BRUTTIG, BICKELL, DOHERTY et al.

anesthesia, as described by Bickell et al.[3,14] These studies were repeated with chloralose/urethane as an anesthetic, and have also been conducted in conscious animals.[15] The results were similar in all cases. Briefly, catheters were placed in various arteries and veins to monitor blood pressures, determine cardiac output, remove blood samples and infuse resuscitative fluids. The site of the uncontrolled hemorrhage was the ventral (belly side) surface of the abdominal aorta, where a 4-0 surgical steel wire was implanted to tear a precise hole in the aorta. Following physiologic stabilization (to allow blood pressure, heart rate and cardiac output to return to baseline levels), the aortotomy hemorrhage was initiated by pulling the wire suture. This resulted in a 5 mm linear defect in the aorta and loss of 35% of the normal blood volume within 3 minutes. Further studies, applying the same type of approach, also have been conducted following venotomy (i.e., a 5-10 mm tear in the vena cava; unpublished data).

Microvascular studies (those conducted on the smallest of the arteries=arterioles and veins=venules, as well as capillaries) were conducted in New Zealand White rabbits (0.8-1.2 kg), which were anesthetized with urethane and prepared for observation of the tenuissimus muscle (lateral thigh) according to descriptions by Borgstrom et al.[16] and Bruttig et al.[17] Surgical preparation was similar except that the wire suture was replaced by a braided 6-0 cardiovascular suture. A microscope was connected to a video camera and recorder, to observe changes in microvascular dimensions and flow. Our objective was to view the smallest of the arteries (arterioles) in the muscle tissue. Microscopic examination showed that terminal arterioles (TE) supply muscle fibers directly, but transverse arterioles (TR), in addition to supplying the terminal arterioles, course through the muscle into the adjacent connective tissue, where they also break up into capillaries. Our microvascular studies were directed at junctions of transverse and terminal arterioles.

Calculations. Microvascular blood flow through the observed portion of the tenuissimus muscle was calculated using a mathematical model developed by Borgstrom et al.[16]

Resuscitation Solutions. Two solutions were used to resuscitate the animals from hemorrhagic hypotension: lactated Ringer's solution (Travenol Laboratories, Inc., Deerfield, IL) and hypertonic/hyperoncotic saline (HSD) solution (7.5% saline/6% Dextran 70, Pharmacia, AB, Uppsala, Sweden).

BRUTTIG, BICKELL, DOHERTY et al.

Immediate Resuscitation. In the systemic studies (whole body, large blood vessel studies in pigs), immediate resuscitation was begun 6 minutes after the hemorrhage. HSD was given as a bolus (rapid, total volume) intravenous injection (4 ml/kg/min) in about one minute. Ringer's lactate was administered over a ten minute period via intravenous injection at 3 times the volume estimated to have been lost during hemorrhage (approximately 2,100 ml of solution). No microvascular studies (rabbits) were conducted following immediate fluid resuscitation.

Delayed Resuscitation. Systemic studies were conducted 30 minutes after hemorrhage. Pigs received HSD solution (4 ml/kg) either as a bolus injection (over 1 minute) or as a slow infusion (over 12 minutes). In the microvascular studies following delayed resuscitation, all rabbits received only the bolus injection. These animals received either HSD or lactated Ringer's solution. HSD was given as above, but was followed with a continuous infusion of Ringer's lactate at 0.25 ml/min/kg for 30 minutes. Ringer's lactate, when used as the primary resuscitation solution, was given as 1-2 ml/min/kg for 30 minutes.

RESULTS AND DISCUSSION

Responses to Aortotomy Hemorrhage (without resuscitation). This is a substantial but survivable hemorrhage (35% of calculated blood volume.[3,14,15] Arterial blood pressure fell rapidly following aortotomy from 90-100 to 25-35 mmHg approximately 2 minutes after the initiation of hemorrhage. By 5 minutes post-hemorrhage, arterial pressure had begun to rise, and continued to recover throughout the course of the spontaneous recovery period (120 minutes). Heart rate increased 17% in the first 10 minutes and eventually rose to 150 beats/minute. Hematocrit (percent red cells in the blood) fell soon after the start of hemorrhage (indicating the reabsorption of tissue fluid into the blood = transcapillary refill) and remained low throughout the spontaneous recovery period.

The output of blood from the heart (cardiac output) fell from 5.0 liters per minute to less than 2.0 liters per minute by 5 minutes after the start of hemorrhage. The resistance to blood flow caused by the constriction of the smallest of arteries distributed throughout the body (= systemic vascular

BRUTTIG, BICKELL, DOHERTY et al.

resistance) fell 50% following the initiation of hemorrhage, as did pulmonary artery pressure, central venous pressure and pulmonary vascular resistance. These findings are contrary to the classic response to controlled hemorrhage which includes a moderate to pronounced rise in pulmonary and systemic vascular resistance.[2]

While these results seem inconsistent with recovery, there is great survival value in this type of response. Blood loss and duration of hemorrhage are directly related to blood pressure. Consequently, the profound decreases in resistance and arterial pressure kept hemorrhage volume to a minimum. In addition, the decrease in blood pressure after hemorrhage causes a re-distribution of tissue fluid to the intravascular compartment (i.e., transcapillary refill). The resulting increase in blood volume causes an increase in cardiac output, dilutes the blood, promotes tissue blood flow, and eventually aids in raising blood pressure.

In the conscious animal, cardiovascular responses to aortotomy hemorrhage were similar to those observed in anesthetized animals, with decreases in cardiac output, blood pressure and peripheral resistance. In addition, the conscious animals exhibited voluntary movements, which cause tissue compression, and which may force stagnant blood back into the general circulation. Furthermore, if observed but left untreated, these animals resume "normal" behavior within 15-20 minutes. They continue to recover over the next 80-90 minutes, and will survive (with no change in behavior) for as long as 4 days -- the longest we have observed post-experiment. Taken together, these changes are effective in improving tissue oxygen delivery and subsequent survival. The only lingering evidence of hemorrhage is the reduced circulating red cell mass, since normal red cell repletion requires weeks to complete.

Microvascular Response to Uncontrolled Hemorrhage. In the tenuissimus muscle, both the transverse (pre-terminal) arteriole and the terminal arteriole constrict following hemorrhage. However, these arterioles release the constriction about 10 minutes after the hemorrhage and continue to recover throughout the remainder of the spontaneous recovery period. Following venotomy, there is virtually no release from the profound arteriolar constriction. The difference in constriction between the two types of arterioles reduces flow to a much greater extent in the muscle fibers _per se_ and effectively re-distributes blood

BRUTTIG, BICKELL, DOHERTY et al.

flow preferrentially to the connective tissue of the muscle. Hematocrit falls (20%) during this same period, indicating that substantial transcapillary refill is taking place at this time. It is unclear whether muscle is the primary site for water reabsorption, but the re-distribution of blood flow to connective tissue effectively diverts the remaining blood to a capillary bed with greater permeability characteristics than skeletal muscle per se.[16,18] This response enhances transcapillary refill even further.

Microvascular observation following hemorrhage also reveals a variety of patterns in arteriolar activity. The most intriguing of these are vasoactivity, vasomotion and arteriolar closure. We define vasoactivity as random, unorganized changes in arteriolar diameter. Vasomotion is the rhythmic, repetitive constriction and dilation of either pre-terminal or terminal arterioles with a regular period (wavelength), either 3-7 seconds or 25-35 seconds. Either of these types of microvascular vasomotive activity is associated with brief periods of vascular narrowing, whereas vascular closure represents protracted periods (in excess of a few seconds) of vascular narrowing.

In general, controlled hemorrhages initiated vasomotion, whereas uncontrolled hemorrhages tended to decrease the incidence of vasomotion, at least transiently. These differences in vasomotive behavior may account for the differences in peripheral vascular resistance (increase vs. decrease) observed following controlled vs. uncontrolled hemorrhages, respectively. These types of microvascular activity have been described both as normal microvascular physiologic behavior[19] and as physiologic responses to abnormal situations.[20,21] While vasomotion may represent a physiologic response to either a physiologic or an abnormal situation, microvascular closure within skeletal muscle fibers represents a profound response, probably to an extremely abnormal situation, and may have dire consequences on muscle tissue metabolism if continued for long periods of time (i.e., allow tissue lactate production to rise to life-threatening levels). In fact, the concern over lactate following hemorrhage has prompted some investigators to develop an equation, based on high plasma lactate levels, to predict survival following massive hemorrhage.[22]

It is remarkable that all of these responses, either by pigs or rabbits, are not usually associated with fatal events,

BRUTTIG, BICKELL, DOHERTY et al.

at least not during the spontaneous recovery period. Both
groups of animals (pigs and rabbits) demonstrated an
unanticipated recovery following this hemorrhage, but the most
robust recovery was observed in the pigs.

Immediate Fluid Volume Resuscitation. Compared to
control animals (pigs), and regardless of the solution used,
those pigs that were resuscitated immediately showed an 82%
mortality. Of the pigs resuscitated with lactated Ringer's
solution, 100% died within 100 minutes of the initiation of
hemorrhage. Of those resuscitated with HSD, 63% died within
the 2 hour post-hemorrhage period. The volume of hemorrhaged
blood within the abdomen was measured at the end of the
experiment. These measurements indicate that when animals
were immediately resuscitated with either HSD or Ringer's
lactate, hemorrhage volumes were doubled or tripled,
respectively.

Vascular pressures and cardiac output fell initially,
characteristic of the type of hemorrhage imposed. With
resuscitation, mean arterial pressure and cardiac output rose
dramatically, due to the large increases in vascular volume,
but these were short-lived increases. After the termination
of the resuscitation, all of these parameters began to fall,
especially for animals resuscitated with lactated Ringer's
solution. It is quite possible that the sudden rise in
vascular volume, cardiac output and arterial pressure put
undue strain on the immature clot forming on the outside wall
of the aorta. As a result, this clot may have eroded, thereby
re-initiating hemorrhage. This type of hemorrhage, followed
by erosion of the clot and re-initiation of hemorrhage, has
been observed in other uncontrolled hemorrhage studies which
utilized peripheral arterial sites.[9,10] This second hemorrhage
may well have led to the events eventually associated with
death in most animals.

Delayed Fluid Volume Resuscitation. All of our rabbit
studies included fluid volume resuscitation begun at 30
minutes post-hemorrhage. Since there were no fatal outcomes
for rabbits who lived to resuscitation, and since early
resuscitation of pigs was usually fatal, we decided to delay
resuscitation in the pig by 30 minutes in an attempt to
improve survival. Furthermore, to distinguish between rapidly
and slowly increasing arterial pressures, we studied the
effects of rapid (bolus) versus slow infusion delivery of the
resuscitative fluid (HSD, 4 ml/kg). When considered without

194

BRUTTIG, BICKELL, DOHERTY et al.

regard for the method of fluid delivery, the observed mortality was 29%, a significant improvement over that mortality seen following immediate resuscitation (82%). When treated as separate groups, 37.5% of the pigs receiving HSD as a bolus died, whereas only 22% of the pigs receiving HSD as a slow infusion died in the two hours following infusion. Forty percent of the control animals died in this experiment as well, indicating that delayed resuscitation, unlike early resuscitation, did not increase mortality. Furthermore, mean arterial pressure did not rise following the bolus infusion of HSD, but the slow infusion was associated with a transient rise in pressure. The rate at which pressure rose was substantially lower in animals receiving delayed resuscitation than it was following immediate resuscitation. As a result, there may have been significantly less strain placed on the clot, with less tendency for erosion and re-initiation of the hemorrhage. This is borne out by the reduced hemorrhage volumes which followed delayed resuscitation.

Computer simulation experiments. The use of a passive computer model of cardiovascular hemodynamics[23] allowed us to estimate values of those variables that cannot be measured reliably. Figure 1A shows a model estimate (dashed line) of the relationship between mean arterial blood pressure and % hemorrhage from actual experiments.[24] Figure 1B shows the estimated systemic vascular resistance (here TPR) of dogs whose cardiovascular reflexes were rendered passive by disruption of the sympathetic nervous system (dashed line). From these measurements, expected changes in TPR due to changes in microvascular radius, or to autoregulation (solid lines) can be calculated. When these changes (solid lines) are summed, they add up to the values estimated by the computer simulation (dashed line). That the predicted curve is approximately the sum of the expected changes in TPR, due to volume-induced changes in blood vessel caliber and to autoregulation,[24] demonstrates the model's validity.

Summary. Certain responses to the uncontrolled hemorrhage model described here are noticeably different from responses previously described for controlled hemorrhage models. The difference in arterial pressure may be due to the greater rate of hemorrhage (the volumes were equivalent). Arterial pressure fell farther and faster with uncontrolled hemorrhage. These results may also be due in part to the unexpected decreases in systemic and pulmonary vascular resistance. We have hypothesized that these changes in

BRUTTIG, BICKELL, DOHERTY et al.

resistance are a function of the combination of blood loss and vascular injury.[15] Arterial hemorrhage (controlled or uncontrolled) caused a transient diversion of blood flow away from skeletal muscle fibers. This diversion lasted 10-20 minutes after the initiation of hemorrhage while constriction was persistent with uncontrolled venous hemorrhage. Vasomotion commonly occurred following controlled hemorrhage, but it did not always occur following uncontrolled bleeding. It is imperative that its role in uncontrolled hemorrhage be resolved.
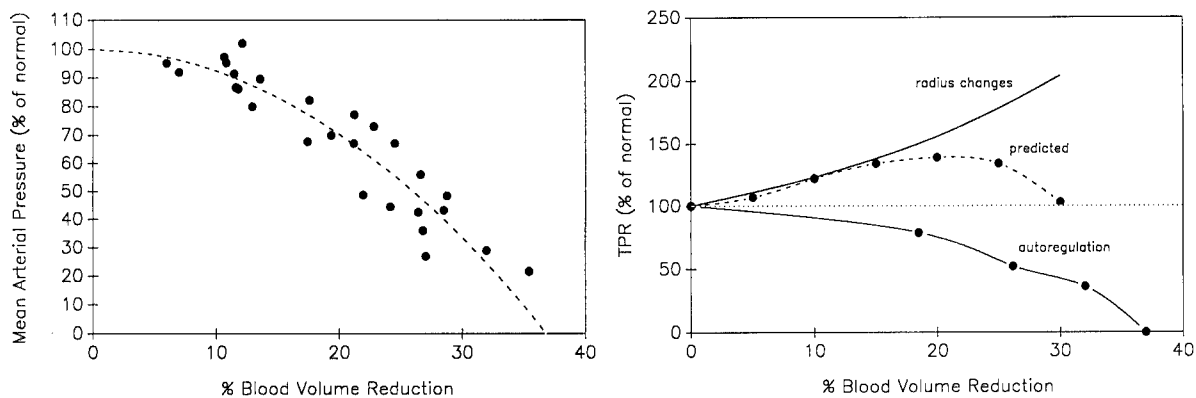
Figure 4. Sample outputs (dashed lines) indicating that the predicted output from the passive model agrees well with observed data. A. Mean arterial blood pressure data, plotted as a function of hemorrhage volume (% blood volume reduction). Filled circles are actual data; dashed line is the computer model estimate. B. Total peripheral resistance (TPR) plotted as a function of hemorrhage volume. Solid lines are calculations of the changes in TPR due to microvascular radius changes or to autoregulation; dashed line is the prediction of TPR from the computer simulation.

BRUTTIG, BICKELL, DOHERTY et al.

The present studies have demonstrated a significant physiologic resilience following the defined aortotomy or venotomy challenges described here, and these studies call attention to the fact that early, aggressive attempts at fluid volume resuscitation of the combat casulty with uncontrolled hemorrhage ("standard of care"), in the absence of adequate control over hemorrhage, may interfere with that resilience. More measured and deliberate approaches, the result of carefully designed animal experimentation, coupled with computer simulation, will take advantage of and hopefully augment that resilience, to maximize survival. Development and/or refinement of our post-hemorrhage resuscitation techniques promise superior health care delivery to the combat casulty, enhanced survival, and a speedier return to duty.

LITERATURE CITATION

1. Hannon, J.P. Hemorrhage and hemorrhagic shock in swine: a review. In: Swine in Biomedical Research, 2nd Edition, ed. by M.M. Swindle, A.R. Liss, Inc, New York, (in press).

2. Chien, S. Role of the sympathetic nervous system in hemorrhage. Physiol. Rev., 47: 214-288, 1967.

3. Bickell, W.H., S.P. Bruttig, G.A. Millnamow, J. O'Benar and C.E. Wade. The detrimental effects of intravenous crystalloid after aortotomy in swine. (In submission to American Journal of Physiology).

4. Caroline, N.L. Emergency Care in the Streets. 2nd ed. Brown, Little and Co., Boston, 1983.

5. Advanced Trauma Life Support Course, Instructor Manual. American College of Surgeons Committee on Trauma, Chicago, 1984.

6. Maningas,P.A., L.R. DeGuzman, F.J. Tillman, C.S. Hinson, K.J. Priegnitz, K.A. Volk and R.F. Bellamy. Small volume infusion of 7.5% NaCl in 6% Dextran 70 for the treatment of severe hemorrhagic shock in swine. Ann. Emerg. Med., 15: 1131-1137, 1986.

7.   Krogh, A.   The Anatomy and Physiology of Capillaries, reprint edition, Hafner Publ. Co., New York, pg. 367, 1959 (original editions 1922, 1929, 1930, Yale University Press).

8.   Guthrie, C.C.   The interpretation of the manifestations of shock.   Penna. Med. J., 22: 123-126, 1918.

9.   Gross, D., E.H. Landau, A. Assalia and M.M. Krausz.   Is hypertonic saline resuscitation safe in "uncontrolled" hemorrhagic shock?   J. Trauma, 28: 751-756, 1988.

10.   Milles, G., C.J. Koucky and H.G. Zacheis.   Experimental uncontrolled arterial hemorrhage.   Surgery, 60: 434-442, 1966.

11.   Zweifach, B.W., R.E. Lee, C. Hyman and R. Chambers. Omental circulation in morphinized dogs subjected to graded hemorrhage.   Ann. Surg., 120: 232-250, 1944.

12.   Berman, H.J. and G.P. Fulton.   The microcirculation as related to shock.   In: Shock and Hypotension, ed by L.C. Mills and J.H. Moyer, Grune and Stratton, New York, pp. 198-219, 1965.

13.   Bohlen, H.G. and R.W. Gore.   Comparison of microvascular pressures and diameters in the innervated and denervated rat intestine.   Microvasc. Res., 14: 251-264, 1977.

14.   Bickell, W.H., S.P. Bruttig, G.A. Millnamow, J. O'Benar and C.E. Wade.   The use of hypertonic saline/dextran vs. lactated Ringer's solution as a resuscitation fluid following uncontrolled aortic hemorrhage in anesthetized swine. Institute Report No. 432, Letterman Army Institute of Research, San Francisco, CA, 1989.

15.   Bickell, W.H., J. O'Benar, S.P. Bruttig, C.E. Wade, J.P. Hannon, F. Tillman and W. Rodkey.   The hemodynamic response to aortotomy in the conscious chronically instrumented swine. The Physiologist, 30:228, 1987.

16.   Borgstrom, P., S.P. Bruttig, L. Lindbom, M. Intaglietta and K.-E. Arfors.   Microvascular responses in rabbit skeletal muscle after fixed volume hemorrhage.   (in Press, American Journal of Physiology).

17.   Bruttig, S.P., P. Borgstrom, K. Snell, M. Biren and K.-E. Arfors.   A comparison of the responses to various types of

hemorrhage in the anesthetized rabbit. XXXI International
Congress of Physiological Sciences, Helsinki, Finland, 9-14
July, 1989.

18.   Smaje, L.H., B.M. Zweifach and M. Intaglietta.
Micropressures and capillary filtration coefficients in single
vessels of the cremaster of the rat. Microvasc. Res., 2: 96-
110, 1970.

19.   Meyer, J.-U., L. Lindbom and M. Intaglietta. Coordinated
oscillations at arteriolar bifurcations in skeletal muscle.
Am. J. Physiol., 22: H568-H573, 1987.

20.   Borgstrom, P., S.P. Bruttig, K.-E. Arfors and M.
Intaglietta. Hemorrhage induced vasomotion in skeletal muscle
of anesthetized rabbits. FASEB J. (in press).

21.   Bruttig, S.P., P. Borgstrom, K.-E. Arfors and M.
Intaglietta. Vasoactivity and vasomotion following
uncontrolled hemorrhage in anesthetized rabbits. FASEB J.
(in press).

22.   Wade, C.E., D.S. Trail, V.L. Gildengorin and J.P. Hannon.
Blood lactate as a prognosticator of survival following
hemorrhage in conscious swine. Lab. Animal Sci., 39: 44-46,
1989.

23.   Doherty, T. and S.P. Bruttig A computer simulation of
cardiovascular hemodynamics and the interrelationships of
factors controlling cardiac output. FASEB J., 3: A693, 1989.

24.   Walker, J.R. and A.C. Guyton. Influence of blood oxygen
saturation on pressure flow curve of dog hindleg. Am. J.
Physiol., 212: 506-509, 1967.

UNCLASSIFIED

BURROWS & NELSON

IV Fluidmaker: Preparation of Sterile Water for Injection
In a Field Setting (U)

*W. Dickinson Burrows, Dr.
James H. Nelson, Dr.
U.S. Army Biomedical Research and Development Laboratory
Fort Detrick, Frederick, MD 21701-5010

INTRODUCTION

The US Army Medical Research and Development Command (MRDC) requires a device to manufacture intravenous (IV) fluids under circumstances of disadvantage, such as a field combat situation, where resupply of medical items is uncertain. The device must produce sterile, pyrogen-free water from a potable source. The product water is to be introduced directly into one-liter bags containing concentrated salts so as to make parenteral solutions suitable for IV infusion into humans. Further requirements for the device are that it be hand-operated, produce a minimum of one liter per hour, and be small enough to be carried as part of the aidman's kit. Presently, the smallest system for production of IV fluids projected to be available to the Army in the field is the Resuscitation Fluids Production System (REFLUPS), a 60 liter/hour device about the size of a console piano with substantial water and power requirements, and employing reverse osmosis (RO) technology. Because a literature review and product search has revealed no devices with the potential to meet the requirements, a developmental initiative is necessary.

The United States Pharmacopeia (USP XXI) defines four categories of water relevant to this study: Purified Water, Water for Injection (WFI), Sterile WFI and Bacteriostatic WFI.[1] The authors perceive that the target for the device in question, (hereinafter the fluidmaker) is Sterile WFI, which must be produced by distillation or RO. In addition, there are rigorous quality criteria for pyrogens and many dissolved materials commonly found in potable water. The one criterion that cannot be met by any fluidmaker (including REFLUPS) at the outset is that the source water satisfy U.S. Environmental Protection Agency (EPA) regulations. Army potable water standards (Table 1) are necessarily much less stringent than EPA standards, which are based on

BURROWS & NELSON

lifetime health considerations for the general population and which
specify maximum contaminant levels for a large number of organic and
inorganic chemicals commonly found in raw water supplies. As an
example of circumstances where EPA criteria cannot be met, potable
water prepared from seawater by means of the 600 gallon/hour reverse
osmosis water purification unit (ROWPU), the most advanced water
treatment system in the Army inventory, has chloride levels exceeding
the EPA secondary standard (250 mg/L) by two-fold or more.

TABLE 1. QUALITY STANDARDS FOR POTABLE WATER

Constituent	U.S. Army Short Term Standard[a] (7 days or less)	U.S. Army Long Term Standard[a] (more than 7 days)	Environmental Protection Agency
Physical			
Color	–	50 units	
Turbidity	Reasonably clear	5 units	
Chemical			
Arsenic	2.0 mg/L	0.2 mg/L	0.05 mg/L
Chloride	–	600 mg/L	250 mg/L
Cyanide	20 mg/L	2 mg/L	
Magnesium	–	150 mg/L	
Sulfate	–	400 mg/L	250 mg/L
Total dissolved solids	–	1500 mg/L	
Chemical Agent			
Hydrogen cyanide	20 mg/L	2.0 mg/L	
Lewisite	2 mg/L	0.2 mg/L	
Mustard	2.0 mg/L	0.2 mg/L	
Nerve agents	0.02 mg/L	--	
Bacteriological			
Coliform	1.0 per 100 mL	1.0 per 100 mL	

a. Reference 2

The fluidmaker concept is based on on RO technology, which is known
to be highly effective in removal of dissolved organic and inorganic
impurities and microorganisms, including viruses. Indeed, potable
water prepared by ROWPU could, in principle, be converted to WFI by
further treatment, as required. However, potable water rapidly loses
its identity through distribution in the field, and it was decided that

BURROWS & NELSON

development of the system should assume anonymity of the potable water supply. An ion exchange unit is additionally required since RO alone can produce WFI according to USP criteria only from high quality potable waters. A water purification filter, such as those used by campers or for end-of-pipe treatment in households, is the third main unit of the system. These filters commonly contain activated carbon, which is effective in trapping organic chemicals not reliably removed by RO, and they provide an independent means of removing pathogens and pyrogens. A membrane filter of 0.2 micron limiting pore size is the final element in the treatment train. Although the water purification filter is expected to produce sterile water, the membrane filter is considered to be an essential backup.

MATERIALS AND METHODS

Challenge Waters

Various water sources were used to challenge the fluidmaker and its individual components. Challenge water sources were tapwater from Fort Detrick and the city of Frederick and water from the Fort Detrick post recreational pond, containing suspended algae. "Worst case water" was prepared by amending tapwater with 900 mg/L of sodium chloride and 500 mg/L of anhydrous sodium sulfate. Synthetic ROWPU water was similarly prepared by amending tapwater with 600 mg/L of sodium chloride.

Reverse Osmosis Units

Two hand-operated RO units were evaluated, both manufactured by Recovery Engineering, Inc., Minneapolis, MN. The Survivor^{TM} 06 (S06), designed for emergency production of drinking water from seawater, is 20 cm long, weighs 1.1 kg, and produces 20 to 25 mL/min from fresh water at a pumping rate of 80 to 100 up-and-down strokes per minute (spm). (This is greater than the pumping rate recommended by the manufacturer, but it is a rate easily maintained indefinitely without fatigue). The Survivor^{TM} A90 (SA90) is 38 cm long disassembled, weighs 3.2 kg, and produces 200 to 300 mL/min from fresh water at a pumping rate of 60 spm. Both units use FilmTec^{R} FT30 spiral-wound, thin-film composite membranes having a sodium chloride rejection of 98% at 225 psi (1.6 M Pa), 25°C and pH 7.

BURROWS & NELSON

Ion Exchange Column

Because a small ion exchange column compatible with the fluidmaker is not available off-the-shelf, one was fashioned from a 15 cm length of PVC tubing of 1.75 cm inner diameter (i.d.). End cap nipples fit 6.25 mm i.d. Tygon[R] tubing. The PVC tube is packed with strong acid-strong base mixed bed ion exchange resin from a fresh Barnstead D8902 cartridge, retained at each end with foam plastic filter material taken from the same cartridge. This resin is reported by the manufacturer to have a capacity of 12.7 equivalents of sodium chloride/ft^3 to produce water with a specific resistivity of 1 Megohm-cm or greater, which meets the USP criterion for WFI.

Water Purification Filter

The First Need[R] water purification filter (General Ecology, Inc, Lionville, PA) used throughout this study is a solid matrix filter containing activated carbon and zeta adsorbants; it is rated at 0.1 micron nominal and 0.4 micron absolute. It is a cylinder 7 cm in diameter and ca. 10 cm long, weighs 0.20 kg and comes with a detachable pump weighing 0.11 kg. Inlet and outlet nipples fit the 6.25 mm inner diameter (i.d.) Tygon[R] tubing used throughout this study. The filter produces 400 to 500 mL/min at the slow pumping rate recommended by the manufacturer. Water holdup is 20 to 30 mL.

Sterilization Filters

For the fluidmaker employing the S06, cellulose acetate syringe filters, 25 mm in diameter and 0.2 micron absolute pore size, are obtained from the Nalge Company, Rochester, NY. This filter is connected by means of 4 mm i.d. Tygon[R] tubing on the upstream side to the First Need[R] filter, and on the downstream side is locked to a REFLUPS concentrate transfer set (a 28 cm long, 4 mm diameter Tygon[R] tube with a spike pin at one end and a Luer-Lok[R] syringe connector at the other), thus providing a means for sterile transfer of the WFI to an IV bag. The 25 mm filters cannot accommodate the higher flow rate of the SA90. For this system a final sterilizing filter of 7.5 cm diameter and 0.22 micron pore size from REFLUPS lot packs is connected on the upstream side by large bore (6.25 mm i.d.) tubing directly to the First Need[R] filter; on the downstream side a rubber stopper and stainless steel adapter effect a connection to the REFLUPS transfer set after removal of the Luer-Lok[R] fitting.

BURROWS & NELSON

Miscellaneous Items

The REFLUPS concentrate transfer set, empty 1 L IV bags, and concentrated Ringer's lactate are all manufactured by Abbott Laboratories, North Chicago, IL, for Sterimatics Company, Bedford, MA, the developer of REFLUPS. Each IV bag has a rubber syringe port, a spike pin port, and an integral 0.8 m tube (4 mm i.d.) for filling with WFI from the REFLUPS system. Adapters for joining 6.25 and 4 mm i.d. tubing are fabricated from stainless steel. Spring scales Model 8004-MO, manufactured by Ohaus Scale Corporation, Florham Park, NJ, have 25 g scale divisions and a capacity of 2000 g. The Super Cub butane torch used for flame sterilization and to seal tubing is distributed by Microflame, Inc., Minneapolis, MN.

Test Procedures

Water treatment components were tested individually and in combination. For the S06, 20 to 50 mL of test water was pumped through the system to flush out residual particles before taking samples or filling IV bags; for the SA90, at least 500 mL was pumped through. Connection to the IV bag was made by means of the concentrate transfer set or through the integral filling tube. IV bags were sealed by heating the filling tube with a butane torch to the point of softening, then immediately pinching off the tube with flat-nose pliers; alternatively, tubes were sealed off temporarily with a slotted plastic tab. For all tests involving the RO unit, source water was first dechlorinated either by allowing an open container of water to stand for several days or by adding 0.1 g of sodium bisulfite to each 5 gal (19 L) of tapwater. RO units were treated with sodium bisulfite while not in use, as recommended by the manufacturer.

Transfer of Concentrate

Individual IV bags containing concentrated Ringer's lactate were prepared by transferring 50 mL of concentrate to each bag by means of a 50 mL syringe with a 5 cm, 13 gauge hypodermic needle. Every effort was made to maintain sterility. The bags were sealed in lots of three in plastic overwraps. Sterility was checked for only a few samples, but it was found that concentrate samples that remained clear after a few weeks were sterile, while non-sterile samples soon developed visible microbial colonies. Sterile transfer was thereby achieved in 12 of 15 samples.

Analytical Methods

Pyrogen content was determined using the Limulus amebocyte lysate (LAL) chromogenic assay. Sterility of water samples was tested by plating on sheep's blood agar and MacConkey's agar. Chloride, sulfate, sodium, potassium, ammonium, calcium and magnesium ions were determined by ion chromatography (Dionex 4000 series, Sunnyvale, CA). Other water quality parameters were determined by procedures given in USP XXI.[1]

RESULTS AND DISCUSSION

Reverse Osmosis

Two hand-operated RO units were investigated for removal of dissolved solids. The Survivor[TM] 06 (S06) is small enough to fit into a protective mask container (a constraint suggested by MRDC); the Survivor[TM] A90 (SA90) is not, but the purified water flux is ten times that of the smaller unit. The principal function of the RO unit is to remove dissolved salts, especially chlorides, sulfates and heavy metals. Polyamide membranes of the type used in the S06 and SA90, as well as in the ROWPU, generally remove 98 percent of dissolved sodium chloride. In practice, both RO units achieved this removal, with the S06 giving superior results. Reverse osmosis will also remove most organic compounds, pyrogens, viruses and other microbes; hence, a device to prepare sterile, pyrogen-free water could in principle be based on reverse osmosis alone. However, experience has shown that the S06 and SA90 are soon colonized by gram negative bacteria downstream from the membranes, thereby necessitating further treatment.

Because Army potable water may have free chlorine residuals exceeding 5 mg/L,[2] it is essential that the RO feed water be dechlorinated to protect the membranes. This can be achieved in the field by addition of sodium bisulfite, which destroys chlorine by the following reaction:

$$2HSO_3^- + HOCl \ ---- \ HS_2O_6^- + Cl^- + H_2O$$
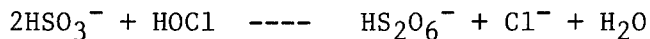
This corresponds to 5.9 mg/L of sodium bisulfite for each mg/L of free available chlorine, or 0.56 g of sodium bisulfite to dechlorinate 5 gallons (19 L) of potable water with an initial 5 mg/L residual. It should be noted that potable water drawn for RO treatment from a finite reservoir, such as a 5-gallon container, will deteriorate in quality as rejected salts build up in the reservoir. In the case of potable water with high levels of dissolved solids, it may be necessary to replenish the source water frequently. The principal disadvantage of batch

BURROWS & NELSON

chemical dechlorination is that the dechlorinated water may need to be discarded and the replenished water dechlorinated anew. Alternatively, a granular carbon filter could be used to dechlorinate the water, but because the SA90 must draw 4 liters and the S06 10 liters for each liter of purified water produced (the reject stream being returned to the source), the carbon filter may be large enough to cause a significant pressure head loss and a consequent reduction in purified water flux, besides adding a bulky component to the system.

Ion Exchange

The small ion exchange unit fabricated for this study has an exchange capacity of 1 g of sodium chloride, presumably sufficient to demineralize 100 L of water containing 10 mg/L of sodium chloride to the specific resistivity of 1 Megohm-cm or greater required for WFI. Although no special effort was made to optimize this unit, it appears to achieve the further purification necessary to meet USP XXI standards for WFI and is sufficiently small. Inclusion of an ion exchange column in the fluidmaker concept is not without complications. Unlike the other components, the ion exchange column must be supported in an upright position; otherwise, flow channeling will occur, and deionization will be incomplete. The ion exchange process must be followed by depyrogenation and sterilization, since ion exchange resins for water purification, as delivered, usually contain adsorbed bacteria, which produce pyrogens. At this stage in development of the fluidmaker, ion exchange technology is considered by the authors to be optional because the product water from RO is safe for IV use in terms of its salt content.

Water Purification Filter

Sterile, pyrogen-free water could be prepared from various water sources using only a First Need[R] solid matrix water purification filter (Table 3). Freshly drawn tapwater had a pyrogen content of 7 to 10 endotoxin units per milliliter (eu/mL); 50 L was passed through a single filter without pyrogen breakthrough. Pond water and other waters with pyrogenicity exceeding 100 eu/mL were tested, but in no instance did pyrogens exceed the detection limit of 0.06 eu/mL in product water from a freshly unwrapped First Need[R] filter, much less the USP XXI standard of 0.25 eu/mL.[1]

TABLE 2. FIRST NEED[R] FILTER PYROGEN TESTS

Sample	Pyrogens, endotoxin units/mL	
	before filtration	after filtration
Frederick municipal water	8.3	0.06
Fort Detrick tapwater	7.3	0.06
Post Pond water	106.1	0.06
Ion exchange water[a]	7.1	0.06
Tapwater, after 1 L	10.2	0.06
" 12 L		0.06
" 25 L		0.06
" 50 L		0.06

a. Water initially pyrogen-free was stirred with an equal volume of ion exchange resin before filtering.

The manufacturer of the First Need[R] filter recommends that the first few ounces of water be discarded when the filter is initially put into use in order to eliminate residual particulate matter. Our experience has demonstrated that these particles will rapidly plug a sterilizing filter, and for this reason it is advisable to flush the system with at least 100 mL of water before the sterilizing filter is installed.

On standing, wetted First Need[R] filters are rapidly colonized by mixed gram-negative bacteria downstream from the filter matrix, thereby compromising both pyrogenicity and sterility. Attempts to restore contaminated First Need[R] filters for reuse met mixed success. Backflushing with chlorinated water, as recommended by the manufacturer, restored sterility, but did not reduce pyrogenicity sufficiently. A combination of disinfection, flushing and autoclaving achieved product water sterility and undetectable pyrogenicity, but one filter was permanently blocked by this process. Restoration of First Need[R] filters for reuse in the field may be feasible, but at this time the only way to assure production of sterile, pyrogen-free water is to discard the filter after each production run, and not attempt to retain the wetted filter for reuse.

BURROWS & NELSON

Sterilizing Filters and Sterile Docking

Although freshly opened First Need[R] filters consistently produce sterile water, these filters are not intended for medical use, and it is considered essential that a sterilizing filter of 0.2 micron pore size follow the First Need[R] filter. For the S06 system, a 25 mm diameter cellulose acetate syringe filter connected to a REFLUPS transfer set permits sterile docking to the IV bag in the laboratory, and should assure sterile docking even under field conditions. The transfer tube can be closed temporarily using a slotted plastic tag or permanently heat-sealed. For the SA90 system, a REFLUPS final filter was used, adapted for the transfer set using a rubber stopper rather than the Luer-Lok[R] connector. This permits sterile docking to the IV bag in the laboratory, but a more satisfactory connection from filter to transfer set will be necessary for field use, particularly since rubber stoppers are known to shed small particles, which are themselves pyrogenic, though undetectable in the LAL test. [A Nylon membrane syringe filter of 0.2 micron pore size (Cameo IV), rated at 100 mL/min, has become available from Micron Separations Inc., Westborough, MA, since completion of the present series of tests. This filter is being used for continued testing of the large RO unit.] With either system, a fresh transfer set must be used for each bag of IV fluid, and the sterilizing filter should be replaced at least as often as the water purification filter.

Product Water

Both S06 and SA90 systems consistently provide sterile, pyrogen-free water; when the ion exchange unit is included, the product water is equivalent in all parameters measured to USP XXI Sterile WFI (Table 3). Samples bagged and stored for seven days (with or without IV concentrate) have remained sterile and pyrogen-free. When the fluidmaker is used to prepare parenterals, it is necessary to provide IV bags containing the parenteral concentrate in either liquid or solid form. For this study, IV bags were prepared containing 50 mL of concentrated (20:1) Ringer's lactate. Unless sealed in plastic overwraps, these bags tend to dehydrate, losing 1 gram of water per bag over a period of 11 days at 85 percent relative humidity and 85°F and about twice as much at 40 percent relative humidity. Food and Drug Administration specifications call for an accuracy of ± 5 percent in finished parenteral concentrations, corresponding to ± 50 mL of WFI per 1 liter bag. Although this level of accuracy can probably be achieved volumetrically, small spring scales, which provided an accuracy of ± 25 g, were used in this study.

BURROWS & NELSON

TABLE 3. PERFORMANCE OF FLUIDMAKER[a,b]

Sample	Cl^- mg/L	$SO_4^=$ mg/L	Na^+ mg/L	Ca^{2+} mg/L	NH_3 mg/L	Pyrogens eu/mL	Sterility[c] colonies/mL
Tapwater	25.5	48	13.9	51.6	NM	123.3	NM
S06 w/o IE	0.26	ND	1.5	ML	NM	0.06	No growth
Worst-case water	580	420	556	56	11	38.1	10,000
S06 w/o IE	1.8	10	1	ML	1	NM	NM
SA90 w/o IE	3.7	31	14	ML	1	NM	NM
S06 with IE	0.1	BD	1	ML	1	0.06	No growth
SA90 with IE	0.06	BD	1	ML	1	0.06	No growth
ROWPU water	405	53	250	41.2	1	NM	10,000
S06 w/o IE	1.3	9	4.4	ML	1	NM	No growth
SA90 w/o IE	3.9	15	11.5	ML	1	NM	No growth
S06 with IE	0.04	BD	1	ML	1	0.06	No growth
SA90 with IE	0.1	BD	1	ML	1	0.06	No growth

a. Complete data are provided in reference 3.
b. IE = ion exchange, NM = not measured, BD = below detection limit, ML = meets limit for calcium according to USP XXI, eu = endotoxin units.
c. Growth 48 hours after plating

SUMMARY AND STATUS OF EFFORT

Two complete IV fluidmaker systems were tested for their ability to provide sterile WFI in 1-liter IV bags from potable water of various quality. Both systems produce sterile, pyrogen-free water equivalent to USP XXI water for injection in all parameters measured. The fluidmaker S06 system, weighing approximately 1.5 kg, meets all requirements and produces WFI at a rate of 1 L in 45 to 50 minutes; the fluidmaker SA90 system, weighing approximately 3.5 kg, is substantially larger, but produces WFI at a rate of 1 liter in 6 to 7 minutes.

At this time, the S06 system is preferred. Besides being smaller and lighter, all components (except ion exchange) are immediately

BURROWS & NELSON

available off-the-shelf. However, in the event that the Survivor^{TM} A90 becomes readily available under the same scenario as projected for use of the fluidmaker, or if its greater size and weight are acceptable, then the SA90 system should be considered favorably for its high production rate. Regardless of the system selected, it is recommended that the ion exchange unit be omitted until a unit suitable for field use can be located or developed. Although the water produced will not then meet USP XXI criteria for Sterile WFI with respect to salts, the authors believe that it will be safe for the intended use.

Efforts to further develop the fluidmaker are proceeding in several directions. An extended series of performance and endurance tests is in progress with fluidmakers based on both the Survivor^{TM} 06 and the Survivor^{TM} 35 (which has replaced the A90 and has a somewhat lower production rate). These fluidmakers have been made more field-ready, principally by devising pressure-secure, quick-release connectors between components. The U.S. Army Biomedical Research and Development Laboratory has initiated an extramural effort to develop an ion exchange module the size of a 12 oz beverage can which can be operated in any position, has an exchange capacity of at least 1 g of sodium chloride, and produces water with specific resistivity of 1 Megohm-cm or better. At this time, devising and validating a system to introduce parenteral concentrates into the WFI present the greatest challenge in fielding the fluidmaker. The authors have learned that the National Aeronautics and Space Administration has undertaken research on the problem of IV concentrates;[4] we shall closely follow progress of this effort.

REFERENCES

1. The United States Pharmacopeia, Twenty-First Revision. 1985. United States Pharmacopeial Convention, Inc., Rockville, MD.

2. US Army. 1986. Occupational and Environmental Health. Sanitary Control and Surveillance of Field Water Supplies. TB MED 577. Headquarters, Dept. of the Army, Washington, DC.

3. Burrows, W.D. and J.H. Nelson. 1988. IV Fluidmaker: Preparation of Sterile Water for Injection in a Field Setting. Technical Report 8814, ADA207411, U.S. Army Biomedical Research and Development Laboratory, Fort Detrick, Frederick, MD.

4. Kenneth Broadwell, MD, NASA Johnson Space Center, personal communication, July 1988.

The Potential of New Army Tests to Improve Job Performance (U)

Dr. Henry H. Busciglio[*], Dr. Jay Silva, & Dr. Clinton Walker
U.S. Army Research Institute
5001 Eisenhower Avenue, Alexandria, VA, 22333-5600

Maximally effective job performance requires high quality in all three of these components: people, training, and leadership. The quality of Army personnel is the focus of this paper, which is an evaluation of new personnel tests for selecting the best persons and assigning them to the right jobs.

Since 1976, the test for assuring the quality of enlisted personnel has been the Armed Services Vocational Aptitude Battery (ASVAB), a measure of general intellectual aptitude. Recently, tests have been developed to measure a number of more specific abilities that are widely believed to be important for success in Army jobs. If these other abilities could be measured at the entry point, then that new information could be used with ASVAB both for screening volunteers and matching their aptitudes with Military Occupational Specialties (MOS) more effectively.

The new tests that are evaluated in this paper -- 10 computerized and 6 pencil-and-paper -- measure many of these specific abilities. Among these are eye-hand coordination (i.e., psychomotor abilities), facility with information on the shape, orientation, direction, and movement of objects in space (i.e., spatial abilities), and aptitude for processing visual information (i.e., perceptual speed and accuracy).

These new tests were developed under the Army's Project A, Improving the Selection, Classification, and Utilization of Army Enlisted Personnel[1]. Although Project A also developed tests of attributes like motivation, adjustment, and discipline, those are not included in the present evaluation. Until now, the Project A ability tests have been analyzed[2] only in ways that did not determine the best combinations of tests nor the benefits of each single test. The present research is the first such evaluation.

This evaluation is in terms of incremental validity. Validity is the consistency with which examinees' scores on the tests correspond with their performance on the job. On a good (i.e., highly valid) test, those who score high (or low) on it tend to perform well (or badly, respectively) on the job. Such a test is said to predict job performance. Validity alone is not enough, however; to pay its own way, a new test must not duplicate ASVAB, but rather add to ASVAB's predictive power. This gain in predictive accuracy is called incremental validity.

213

In this project, the specific research questions were: 1) How much can optimal combinations of the new tests add to ASVAB? 2) Is either type of new test, pencil-and-paper spatial or computerized perceptual/motor, generally more useful in supplementing ASVAB? 3) Which specific new tests make the greatest contribution to these gains? The first question asks for a best case estimate of the gain to be had from the new tests. The second question bears on whether the computerized tests have a unique contribution to make, or if the gains can be realized without the expense of computerizing testing. The third question seeks the specific new tests that are most likely to realize gains.

This effort took two paths to evaluating the Project A tests: with respect to measures of overall job performance (i.e., "the whole job"), and with respect to specific combat critical tasks (viz., gunnery). These two perspectives on performance, which are discussed in turn below, were examined because of their value to the Army, the first for its breadth and the second for its high impact.

GAINS IN PREDICTING OVERALL JOB PERFORMANCE

Examinees

Examinees were over 4,039 first tour soldiers in nine MOS (Table 1). These MOS had been selected for their criticality and for their representativeness in terms of Army tasks, personnel, and ASVAB classification scores used. The new tests were administered at 13 posts in CONUS and USAREUR to soldiers who were 18 to 28 months into their first tour.

Table 1
Subjects for the Present Analysis

| MOS | Enlisted Job | Measures | |
		Predictors	SQT
11B	Infantry	491	444
13B	Cannon Crew	464	396
19E	Armor Crew	394	338
31C	Single Channel Radio Operator	289	248
63B	Light Wheel Vehicle Mechanic	478	409
64C	Motor Transport Operator	507	427
71L	Administrative Specialist	427	361
91A	Medical Specialist	392	0
95B	Military Police	597	545
TOTAL		4,039	3,168

Predictor Tests

Table 2 lists the predictor tests which were analyzed. Scores for the separate subtests of ASVAB were used. All of the six spatial and ten computerized tests from the Project A battery were included. Separate speed and accuracy scores were included for four of the computerized tests.

Table 2
Predictors Used in the Analysis

ASVAB Subtests:	Spatial Ability Tests:
Mechanical Comprehension	Assembling Objects
Auto/Shop Information	Map
Electronics Information	Maze
Math Knowledge	Object Rotation
Arithmetic Reasoning	Orientation
General Science	Figural Reasoning
Coding Speed	
Number Operations	
Verbal (Paragraph Comprehension + Word Knowledge)	

Computerized Tests and Scores:
  Psychomotor
      One-hand Tracking   - accuracy
      Two-hand Tracking   - accuracy
      Target Shoot   - accuracy and time-to-fire
      Cannon Shoot   - time discrepancy (from optimal)
  Perceptual
      Simple Reaction Time   - decision time
      Choice Reaction Time   - decision time
      Perceptual Speed and Accuracy - decision time and proportion correct
      Target Identification   - decision time and proportion correct
  Memory
      Short-Term Memory   - decision time and proportion correct
      Number Memory   - response time

Performance Measures (Criteria)

Five measures of overall job success were used, four of these from Project A. The four were scores on Knowledge Tests, Hands-on Tests, General Soldiering Proficiency, and Job-specific Proficiency[3]. The Knowledge Tests were in traditional paper-and-pencil multiple-choice format and covered 25 to 31 critical tasks. For these analyses, scores on both end-of-training knowledge (130 - 210 items) and job knowledge (150 - 200 items) tests were combined into a total Knowledge Test score.

The Hands-on Tests were job samples (i.e., tests of soldiers' skill in carrying out tasks for their jobs). Senior Non-commissioned Officers observed and scored these tests. For each MOS, from 14 to 17 tasks were tested. Individual steps in the tasks were graded pass/fail, and the number of steps passed per task was recorded. General Soldiering Proficiency consisted of tasks common to all MOS, such as first aid, navigation, and target identification, whereas Job-Specific Proficiency consisted of tasks that are shared by few, if any, other MOS. The two measurement methods were crossed with the two types of content, so Knowledge and Hands-on Tests each covered both common and job-specific tasks. Similarly, each of the proficiencies was covered in both written and hands-on modes.

The one measure of overall job success that did not come from Project A was the Skill Qualification Test (SQT), a written multiple choice test developed by the U.S. Army Training and Doctrine Command. It is used periodically to evaluate soldiers' proficiency in their MOS. For MOS 91A, SQT scores were not available.

The present research did not emphasize measures of performance that are in use in schools and units, because such measures are not uniformly developed, used, or recorded. For example, scores are often missing and often are merely pass/fail. Thus, their lack of specificity and completeness makes them hard to interpret.

Procedure

Measures. The examinees were tested on Project A predictors and performance measures at their home posts in 1985. All testing was done with the understanding that the scores would be used for research purposes only. For each soldier, the testing took place over two consecutive days, the sequence of tests differing across soldiers to allow simultaneous testing of as many examinees as possible. Pre-enlistment ASVAB test scores were used for these analyses. Thus, ASVAB scores had been generated at least 18 months earlier, when these same people, as applicants, were qualifying for enlistment.

Analyses. Validity and incremental validity were determined by a series of regression analyses. Before the validity of the individual ASVAB subtests was determined, the examinees' ASVAB scores were statistically corrected for restriction of range. ASVAB's predictive power tends to be underestimated when the uncorrected scores of soldiers are used, because applicants who score very low on ASVAB are screened out before entry. To correct for this narrowing in ASVAB scores, matrices of estimated population (i.e., unrestricted) covariances[4,5] among predictors and criteria were estimated using the Lawley formula[6]. Then, these matrices were used as input for the regressions. Finally, validities were adjusted for shrinkage on cross validation using Wherry's method[7].

Backward stepwise regressions[8] were run on the corrected ASVAB scores and scores on the Project A tests separately for each performance measure. In the first stage of the regressions, all subtests of ASVAB were optimally weighted in an equation for predicting a given criterion. Next, each subtest was tested in turn for removal from the equation. Subtests not contributing significantly ($p > .05$) were removed, and the process continued iteratively until only significant predictors remained.

To the retained ASVAB subtests, the Project A tests were added in two blocks: Spatial and Computerized. For each block, the backward elimination procedure was repeated and the significant contributors were retained. This procedure was run twice for each performance measure, once with the Spatial Tests entered first and once with Computerized first. Thus, retained ASVAB subtests were the baseline for testing incremental validity of second stage tests, and ASVAB plus the retained second stage tests were the baseline for the third stage.

As an indication of the relative contribution of the individual tests[9], semi-partial correlations were computed for each predictor at Stage 3 of the analyses. These were computed for every combination of criterion, MOS, and order of entry of the blocks of new tests. For each predictor, the square of the semipartial is the proportion of total variance in that performance measure uniquely associated with it.

RESULTS

Table 3 shows the results of the regression analyses at each stage, giving the squared multiple correlation (R^2, proportion of criterion variance accounted for) for ASVAB subtests alone, for one set of new Project A tests, and finally for the three together. Both orders of entry for the Project A tests are given. Results are given for each criterion, order of entry, and MOS.

Incremental Validities of Sets of New Tests

In Table 3, for Knowledge Tests, General Soldiering, and Job-specific Proficiency, optimal combinations of the new tests increase the proportion of variance in performance accounted for by ASVAB alone by an average of 7 to 9%. In the case of Hands-on Tests, the gain in R^2 from .29 to .33 is more than a 13% improvement, whereas the gain in predicting SQT is only about 4%. The entries in Table 3 are medians across MOS, so they are not isolated results.

Table 3
Percent of Variance in Performance Accounted for by Significant Predictors
(Median Values Across MOS)

Stage Predictors:	(1) ASVAB	(2a) (1)+SP	(3a) (2a)+P/M	(2b) (1)+P/M	(3b) (2b)+SP
Total Score on School and Job Knowledge Tests:					
MEDIAN	.59	.64	.65	.61	.65
Total Score on Hands-On Tests:					
MEDIAN	.29	.33	.33	.31	.33
General Soldiering Proficiency:					
MEDIAN	.47	.51	.53	.50	.53
Core Technical Proficiency:					
MEDIAN	.44	.49	.50	.48	.51
Skill Qualification Test Score:					
MEDIAN	.53	.54	.55	.54	.55

Note. ASV = Significant ASVAB subtests (all retained). (1)+SP = Retained ASVAB and significant Spatial Tests. (2a)+P/M = Retained ASVAB and Spatial Tests, plus significant Computerized Tests. (1)+P/M = Retained ASVAB and significant Computerized Tests. (2b)+SP = Retained ASVAB and Computerized Tests, plus significant Spatial Tests.

Relative Increment of Spatial and Computerized Tests

The difference in increment produced by the new Spatial and Computerized Tests favors the Spatial Tests by a small margin, as seen in Columns (2) of Table 3. The common medium of testing does not consistently favor the Spatial Tests in predicting written criteria: Spatial Tests add a median of 5% in predicting performance on Project A Knowledge Tests compared with a 2% gain for the Computerized Tests. In contrast, both types account for only an additional 1% of variance in SQT, which is also a written test.

Incremental Validities of Individual New Tests

Table 4 shows median (across MOS) semi-partial correlations for the ASVAB and Project A predictors. These medians were determined from the semi-partial correlations at Stage 3 for the predictors which were significant at any stage in the analyses. The largest values for a new test, .15 against Hands-on and General Soldiering Proficiency, are for the Two-hand Tracking test. On the other hand, the most consistent gains across criteria come from % correct on Short-term Memory and three Spatial Tests. The best of those, Assembling Objects and Figural Reasoning, added significantly to ASVAB more often than six of the ASVAB subtests entered initially. Here again, one medium of predictor testing was not consistently better.

GAINS IN PREDICTING GUNNERY PERFORMANCE

The next evaluation of new tests from Project A relies on specific measures of gunnery performance. As criteria for evaluating new predictors of job performance, combat tasks are revealing because of their criticality. If a new test were to predict gunnery performance, then it could improve the Army's combat effectiveness when used to place persons with the right aptitudes in gunners' position.

The data came from two testbeds of Project A tests. Students in the courses for 11H Anti-tank Gunners and 19K Armor Crewmen were the examinees for this evaluation. In both samples, the students took the Project A tests before entering Basic Combat Training in 1988 under instructions that the test counted and would be recorded in students' records. Administrative obstacles made it impossible, however, for the test scores to influence students' assignments. Near the end of Advanced Individual (i.e., entry-level technical) Training the same students were tested on high fidelity simulators of their respective weapon systems. Two Computerized Tests, One- and Two-hand Tracking, and two Spatial Tests, Orientation and Mazes (Table 2), were the predictor battery for the testbeds. Incremental validity was evaluated by reference to the students' pre-enlistment scores on the subtests of the ASVAB.

11H Anti-tank Gunners

Students qualify as 11H TOW anti-tank gunners on a high fidelity simulator. It requires the student optically to track a target moving across the line of sight on the gunnery range, which is level. The gunner tries to keep a crosshair on the target

with the weapon's hand controls. A sensor in the simulator records the amount of time that the weapon detects an infrared signal from an emitter on the target.

Table 4
Stage 3 Semi-partial Correlations; Medians Across MOS at Stage 3 for Predictors of Comprehensive Performance Measures That Were Significant at Any Stage

	Knowledge	Hands-on	General Soldiering	MOS Specific	SQT
ASVAB:					
Math Knowledge	.07	.09	.08	.06	.07
Auto/Shop	.13	.14	.12	.14	.11
Mechanical Comp.	.01	.08	.05	.04	.05
General Science	.12	.07	.09	.10	.08
Verbal	.08	-.10	.10	.10	.13
Arithmetic Reas.	.05	-.01	.04	.05	.10
Coding Speed	.06	.15	.03	.04	ns
Number Operations	.07	.13	.08	.12	.12
Electronics Info.	.08	ns	ns	.08	.07
Spatial Ability:					
Assembling Objects	.11	.06	.11	.09	.11
Figural Reasoning	.07	.12	.09	.09	.08
Map	.14	.07	.11	.10	.10
Object Rotation	ns	.12	.09	.10	-.06
Orientation	.08	.09	.06	ns	.08
Maze	.05	.07	ns	-.07	ns
Computerized:					
Target Ident % corr	.07	.10	.08	.08	.07
Target ID decis time	.03	.14	.04	.05	.08
Short Term Memory % corr	.11	.08	.11	.07	.14
Number Mem resp time	.07	.07	.07	.07	ns
2-hand Track'g dist.	.07	.15	.15	.08	.05
Percep Spd/Acc time	-.08	-.13	-.08	-.01	-.08
1-hand Track dist.	-.07	-.03	.06	-.08	.10
Cannon Shoot time	-.07	.11	.08	-.09	ns
Short Term Memory decision time	ns	.11	.10	.06	.08
Percep Spd/Acc % corr	.05	ns	.09	.05	.11
Choice Reaction Time decision time	.06	-.11	ns	.10	ns
Target Shoot time	-.07	ns	ns	-.07	-.09
Target Shoot dist	ns	ns	.04	ns	-.08
Simple Reaction Time decision time	ns	.10	ns	ns	ns

The qualifying trials are scored in sets of 10 target engagements. Each engagement is scored for accuracy in terms of percent of time the crosshair is on target. Then, the score for the set is simply the sum of the scores for the ten

engagements. The first two sets are for training only, and the additional sets count in qualifying for the MOS. If a student does not attain a score of 550 by the eighth set, he does not qualify for the MOS. On attaining the qualifying score, the student fires no more sets. Two measures of performance were used: 1) score on first qualifying set, and 2) pass or fail on that set. The first reflects accuracy with practice held constant, the second the need for additional gunnery trials, which is a rough indicator of cost.

Results and Discussion

Table 5 shows the zero-order validities of the ASVAB subtests and Project A tests against the two criteria. A number of ASVAB subtests were significantly related to either or both pass/fail or score on the first qualifying set. In addition to the single subtests, the operational composite that all Services use for selecting applicants, the Armed Forces Qualification Test (AFQT), significantly predicted performance on both of these criteria.

Turning to the new tests, only the psychomotor tests (i.e., both One- and Two-hand Tracking) significantly ($p < .05$) predicted performance on both criteria. Two-hand Tracking fared somewhat better for both criteria, although the difference between the correlations of the two was not significant ($p > .05$). The cause for this pattern of results may be that Two-hand Tracking test has a higher retest reliability than One-hand Tracking does[10]. That advantage could be an artifact of the order of testing; Two-hand Tracking has always been administered after One-Hand.

The regression method described above for evaluating incremental validity was used here. The only change was that logistic regression was substituted for least squares regression when gunnery was scored pass/fail. The new tests added a significant ($p < .05$) amount of validity to ASVAB subtests. For the score on the first qualifying set, 2-Hand Tracking and Mazes predicted an additional 9.6 percent of the criterion variance, which more than doubled the variance in performance that is associated with scores on ASVAB alone. Since Mazes alone did not predict the gunnery scores, its joint impact came about by suppressing variance in the other tests.

Better scores on Two-hand Tracking were also significantly correlated with passing on the first qualifying trial. In terms of Somer's D_{yx} (i.e., an index of rank correlation between predicted probabilities and observed outcomes), adding 2-Hand Tracking to ASVAB raised the index from .292 to .486, for a gain of .194.

In summary, Two-Hand Tracking had a high and consistent validity for predicting TOW gunnery performance. Although it was valid in predicting both outcome measures by itself, One-hand Tracking did not add significantly to the other predictors. Mazes had a significant relation with gunnery performance, but only by removing non-criterion related variance from the other predictors.

Table 5
Validities of Project A Predictors Against Anti-Tank Gunnery

| | Gunnery Measure | |
Predictor	Pass/Fail at First Qualifying Set	Score on First Qualifying Set
ASVAB		
Arithmetic Reasoning (AR)	.17[*]	.17[*]
Auto and Shop (AS)	.10	.16[*]
Coding Speed (CS)	.04	.08
Electronic Info (EI)	.03	.10
General Science (GS)	.12	.15[*]
Mechanical Comprehension (MC)	.07	.21[*]
Mathematical Knowledge (MK)	.08	.13
Numerical Operations (NO)	.04	.00
Verbal (VE)	.08	.10
Armed Forces Qualification		
Test (AR + NO + 2VE)	.14[*]	.16[*]
Spatial		
Mazes	-.05	-.03
Orientation	.03	.13
Psychomotor[1]		
1-Hand Tracking	-.18[*]	-.28[*]
2-Hand Tracking	-.25[*]	-.33[*]

Note. N=229 in 1988. [*] $p < .05$. [1]Good performance produces lower scores on the Tracking Tests and higher scores on the gunnery measures. Hence the signs of those correlations are negative.

M1 Armor Crewman

Students are trained and tested in tank gunnery on the Institutional Conduct of Fire Trainer (ICOFT), a high-fidelity tank gunnery simulator. In this research, students were tested in their 20th hour in the ICOFT, firing 35 engagements in a mix of clear and degraded modes of operation[11]. Both speed and accuracy of firing are recorded. The measure of performance used in the analyses is a composite formed by subtracting the standardized opening time (i.e., time from when a target appeared until the first round is fired) from the standardized percent hits.

Results

Table 6 presents all zero-order validities for the ASVAB and spatial/psychomotor predictors. Many of the ASVAB subtests that predicted scores on the TOW did also

on the ICOFT. The two highest validities for ASVAB were for the Auto and Shop and the Mechanical Comprehension subtests (.33 and .42, respectively). All of the Project A predictors in this testbed significantly related to the speed/accuracy score on ICOFT, Mazes and Orientation to the same degree (.35). The Tracking tests were also related to the criterion to about the same extent (i.e., -0.40 and -0.43).

Table 6
Validities of Project A Predictors Against Tank Gunnery

Predictor	r
ASVAB	
Arithmetic Reasoning (AR)	.21*
Auto and Shop (AS)	.33*
Coding Speed (CS)	.06
Electronic Information (EI)	.03
General Science (GS)	.24*
Mechanical Comprehension (MC)	.42*
Mathematical Knowledge (MK)	.18*
Numerical Operations (NO)	-.06
Verbal	.23*
Armed Forces Qualification Test (2VE + NO + AR)	.23*
Spatial	
Mazes	.35*
Orientation	.35*
Psychomotor[1]	
1-Hand Tracking	-.40*
2-Hand Tracking	-.43*

Note. N=244 in 1988. *p < .05. [1]Good performance produces lower scores on the Tracking Tests and higher scores on the gunnery measures. Hence the signs of those correlation are negative.

To test the predictive power of combinations of the tests, backward regressions were again used. In addition to ASVAB, Two-Hand Tracking and Mazes accounted for an additional 10.5 percent of the variance in gunnery scores. That amounts to over a 52 percent increase in explained criterion variance. Although Orientation and Mazes had equal zero-order correlations with the speed/accuracy score, Mazes shared more variance with the performance measure once both the ASVAB subtests and Two-Hand Tracking were included.

In summary, the usefulness of Two-Hand Tracking was again strongly supported. Mazes also added predictive power, and for this criterion it added it through a direct

relationship rather than by removing non-criterion related variance from the other predictors.

DISCUSSION

Even though ASVAB itself is a powerful predictor of both overall job performance and gunnery, new Project A tests add significant power. The overall performance measures cover content that is so diverse (e.g., Knowledge Tests) and of such universal importance (e.g., General Soldiering Proficiency) that the gains from the contribution of the new tests is shown to generalize widely. We conclude that implementation of the new tests would produce gains in performance that would be large and general across MOS. On specific gunnery tasks as well, the gains in performance to be had by selecting gunners with special aptitudes would be evident in terms of combat effectiveness (higher hit rates) and dollars (e.g., fewer TOW missiles @ $12,000 expended).

As to which type of test has the best potential for improving job performance, the Spatial Tests did slightly better against measures of overall performance, and two of them, Figural Reasoning and Assembling Objects, were most consistent in producing those gains. In contrast, the Tracking Tests did much better against the performance in tank and anti-tank gunnery. Although the Computerized Tests showed some promise against overall job performance and Spatial Tests had some success in predicting gunnery, those results were not as strong or consistent. This pattern of results suggests that the Spatial Tests could be useful in general screening, while the Tracking Tests can increase the Army's effectiveness in matching the abilities of qualified applicants with specific combat critical jobs. The choice of new tests to implement could be driven in part by the aspects or measures of performance that have the highest value to the Army.

One methodological caution is needed regarding the analyses of overall job performance: some aspects of the design favored higher validities for ASVAB while other aspects favored the new predictors. Favoring ASVAB, the examinees in Project A had been tested as applicants who were trying to qualify for enlistment; they were probably trying to do their best. Also, ASVAB has been refined through many cycles of development. The Project A tests had undergone only one round of revision previously, and examinees knew that the scores were not to be used for decisions about their employment. On the other hand, the Project A predictor and performance tests were taken concurrently, so a number of extraneous influences, such as individuals' health, morale, and training could affect scores on the predictors and performance measures similarly. ASVAB scores, which had been generated at least 18 months earlier, may have come under a different set of extraneous influences, which would have reduced their correlation with the same examinees' scores on the later performance tests.

The net effect of these various possible contaminants of the test scores is not clear. The results of the testbeds on gunnery performance encourage us to estimate that it is small. In the testbeds, a predictive design was used, with the Project A tests being taken on entry into the Army and under instructions that the test scores

would count. So with the possible sources of irrelevant variance in Project A reduced, the testbeds still yielded positive results.

Revisiting the question of the generalizability of these results, repeating the same analyses across nine MOS, and using performance measures that overlapped (e.g., Hands-on Test scores were a part of both General Soldiering and MOS-specific Proficiency), there were many opportunities for predictors to attain statistical significance by chance alone. Surely, some of the results presented above are chance findings. However, ASVAB was analyzed under similar conditions of redundancy, so comparisons with ASVAB did not bias the chance findings to favor the new tests.

In the case of gunnery, the results here show that Tracking Tests have a validity that generalizes over two weapons systems in separate samples of examinees. But that replication does not indicate that Tracking is needed for all weapons systems: other research has failed to find validity of the four-test gunnery aptitude battery against performance on high fidelity simulators of the Stinger and Chapparal air defense weapons systems.[16] Rational analysis suggests that the homing feature of those two weapons reduces the need for tracking skill in the operators.

Two findings encourage care in dealing with gunnery data like those above. First, some of the results for anti-tank gunnery vary with the specific outcome measure. That finding suggests that single outcome measures will not provide as firm a basis for evaluating tests as would a richer body of performance data. In connection with this point, note that the measure of tank gunnery used here was a composite of speed and accuracy.

Second, the differences across tank and anti-tank gunnery in validities of both ASVAB and the new tests may show that somewhat different skills are needed for success on the two systems. The tank gunner is more isolated from the outside environment and thus may rely more on the spatial visualization and orientation skills measured by the Mazes and Orientation tests. Here again, such specialized ability tests show promise for improving the person-job match. Task analyses of other gunnery jobs may suggest which aspects of the different weapons systems demand which specific skills.

This research shows that new tests from Project A can markedly improve the quality of Army enlisted personnel. With good training and leadership, the soldiers that are selected and assigned to jobs with these tests can preserve the effectiveness of the Army.

REFERENCES

1. Peterson, N. G. (Ed.). (1987). Development and field test of the trial battery for Project A (ARI Technical Report 739). Alexandria, VA: U.S. Army Research Institute.

2. Campbell, J. P. (Ed.). (1988). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1986 fiscal year (ARI Technical Report 792). Alexandria, VA: U.S. Army Research Institute.

3. Campbell, ibid.

4. Department of Defense. (1982). Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery. Washington, D. C.: Office of the Assistant Secretary of Defense (Manpower, Reserve, Affairs, and Logistics).

5. Palmer, P., Hartke, D. D., Ree, M. J., Welsh, J. R., & Valentine, L. D. (1988). Armed Forces Vocational Aptitude Battery (ASVAB): Alternate forms reliability (Forms 8, 9, 10, and 11) (Report AFHRL-TP-87-48). San Antonio, TX: Air Force Human Resources Laboratory.

6. Lord, P., & Novick, M. (1968). Statistical theory of mental test scores. Reading, MA: Addison-Wesley Publishing Co.

7. Wherry, R. J. (1940), Appendix A. In W. H. Stead and C. P. Shartle (Eds.), Occupational counseling techniques. New York: American Book Company.

8. Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

9. Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd. Ed.). New York, NY: Holt, Rinehart and Winston.

10. Toquam, J., Peterson, N., Rosse, R., Ashworth, S., Hanson, M. A., & Hallam, G. (March, 1986). Concurrent validity data analysis: Cognitive paper-and-pencil and computer-administered predictors (Trial Battery). Unpublished data analyses by Personnel Decisions Research Institute, Minneapolis, Minnesota.

11. Graham, S. E. (1989). Assessing the impact of psychomotor and leadership selection tests on the Excellence in Armor Program (Research Report 1510). Alexandria, VA: Army Research Institute.

12. Gast, I. F., & Johnson, D. M. (November, 1988). Evaluating psychomotor and spatial tests for selecting air defense gunners. Paper presented at the 30th Annual Conference of the Military Testing Association. Arlington, VA.

TACTICAL SOURCE REGION ELECTROMAGNETIC PULSE SIMULATION (U)

M. Bushell, R. Fleetwood, G. Merkel, M. Smith, G.W. Still
Harry Diamond Laboratories
Adelphi, Maryland

Introduction:

The source terms that drive the electromagnetic field in the tactical source region (TSR) are provided by high-energy radiation from a nuclear environment through the mechanism of the Compton effect (which leads to space current and charge separation) and by current and charge density induced in local or distant objects (including the ground). The region in which the Compton current, $j(x,t)$, and charge density, $\rho(x,t)$, are nonzero is called the source region, and the term source-region EMP (SREMP) is used whenever these local drivers are significant.[1,2] In a typical tactical environment, sources throughout the region contribute, through the well-known mechanisms of electromagnetic radiation and diffusion to the fields at any given point.

Because of the extent of the source region, its size lies at the heart of the greatest challenge to successful SREMP simulation. Powerful sources of high-energy pulsed radiation (e.g., Aurora[3] and Hermes II[4]) are available, and these naturally produce Compton current and charge separation in the air into which the radiation is released. There is a temptation to assume that if these local sources are reasonably well reproduced, the full electromagnetic environment will be also. This naive assumption neglects the contribution of "distant" sources (because of the size of the source region) that cannot be reproduced, since existing pulsed-radiation technology can irradiate a volume whose linear dimensions are expressed in tens of meters at most (for example, the Aurora test cell measures $5 \times 12 \times 20$ m³).

Of course, the "prompt" environment can be reliably reproduced because of the finite velocity of light. However, the absence of sources more than 5 m from the field point becomes noticeable after 10 ns or so. In addition, this effect is aggravated in simulation testing at Aurora by the electromagnetic boundary conditions imposed at the metal walls bounding the test cell which short out the electric field.

In this paper, we will briefly discuss the progress of the SREMP simulation research carried out at Harry Diamond Laboratories (HDL) over the past several years. We will specifically discuss the possibility of simulating the tactical source region EMP with a combined e-beam, bremsstrahlung environment.

Review of HDL SREMP Work:

In the Mark I SREMP simulator,[2] shown in Figure 1, an EM field was produced by a parallel plate transmission line driven by a 100-kV pulser. The idea of this simulation was
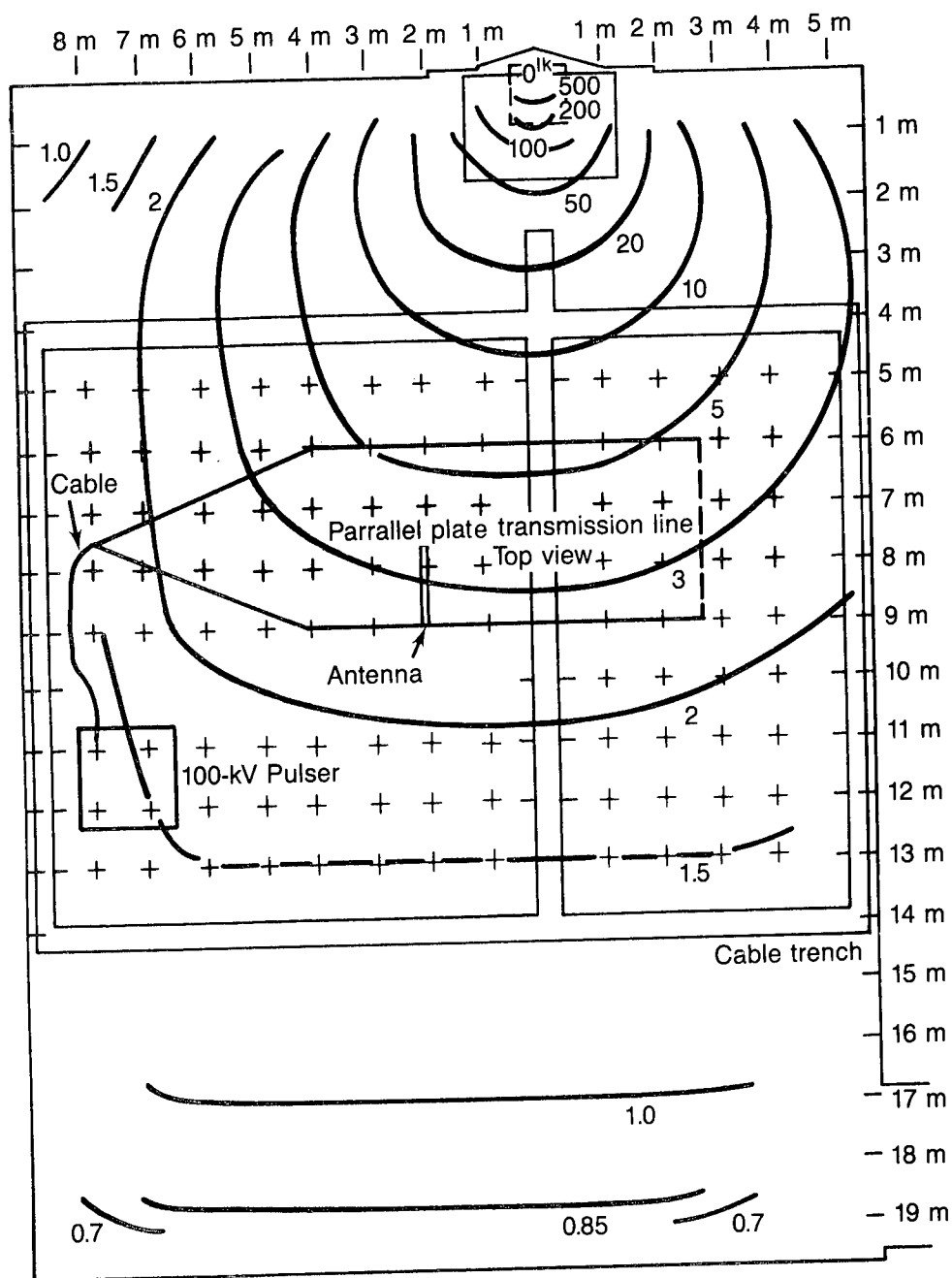
Figure 1. Schematic top view of transmission line in Aurora test cell

essentially that proposed by Longmire,[5] except that the transmission line was driven by the pulser. It produced a 30-kV TEM wave between the plates of the transmission line. The purpose of the Aurora flash x-ray machine in the Mark I SREMP simulator experiments was to provide bremsstrahlung that produced Compton electrons to ionize the air.[6-9] In this transmission line technique, only a very small part of the electromagnetic field was produced by Compton electrons; over 90 percent was produced by the parallel plate transmission line. Figures 1 and 2 show this experimental arrangement. The sole purpose of the Aurora ionizing radiation was to generate time-varying air conductivity. The transmission line produced an electromagnetic field. The Aurora and transmission line were fired simultaneously. This technique and possible improvements of this scheme are discussed in much greater detail elsewhere.[7,10] In the work presented in this paper we examine the possibility of replacing the auxiliary transmission line with an e-beam.

The Hermes II SREMP simulation experiments[4] (carried out by workers at HDL and Sandia National Laboratory) also investigated the possibility of replacing an auxiliary electromagnetic field source with an e-beam. These experiments at Hermes II had the advantage that they were carried out in the open air, hence the electric fields generated were not shorted out by the test cell walls as they were in the Aurora test chamber (Fig. 2). Experimental results demonstrated that a proper TSR-like environment could be produced.[4]

STUDY OF RELATIONSHIP BETWEEN E-BEAM PULSE SHAPE AND RESULTING AURORA TEST CELL ELECTRIC FIELD

Typical electric field waveforms generated with the modified e-beam at Aurora are shown in Figures 3 and 4 for several different sensor configurations. Figure 3 compares two electric field measurements obtained at the same location in the test cell using two parallel plate sensors at differing separations but with the same 10-kohm series resistance. Figure 4 shows a comparison of two measurements with the same sensor separation but with differing series resistances of 10 and 20-kohm. These measurements were done to ensure that the sensor was responding to fields rather than noise. Figure 5 gives a comparison of measured and computed electric fields on the Aurora test cell floor. An unaltered conventional e-beam pulse was injected into the test cell for these measurements.[8,9,11] The series of measurements shown in Figures 3 and 4 and the comparison to calculation shown in Figure 5 are consistent and show a common wave form. These electric fields can be divided into three time phases. The first fast rising phase is produced before ionized air conductivity builds up in the test cell. The second phase (corresponding to the low E-field valley) occurs when air conductivity shorts out the E-field. The final slowly decaying after-pulse results from the fields produced by an excess of negatively charged O_2 ions.[9,11] This excess of O_2 ions is formed when relativistic electrons have trajectories that terminate in the volume of the test cell. Some of these electrons attach to O_2 to produce O_2 negative ions. When the e-beam pulse ceases, air conductivity becomes ionic. The excess negative charge relaxes by means of ionic conductivity.[9,11]

Electron beam nose erosion (drift tube nose erosion) combined with late time beam clipping (crowbarring by variable diverter setting) can be employed to vary the Aurora e-beam pulse shape.[11-14] Electron beam parameters (risetime, falltime, beam current magnitude, and electron energy distribution) were varied in order to determine if the electric field waveform produced in the Aurora test cell could be controlled by a combination of e-beam risetime enhancement (using drift tube nose erosion) and e-beam pulse length clipping (by adjusting the diverter setting). The ultimate goal was to see whether the electric field in the test cell could be made
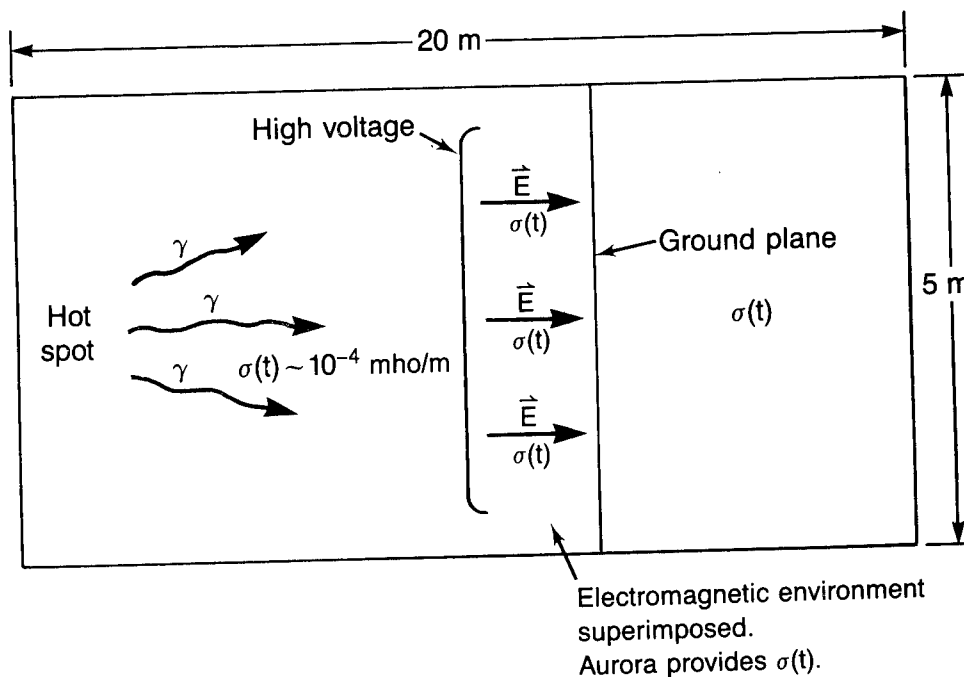
Figure 2. Schematic view of a transmission line designed to produce an auxiliary electric field in the Aurora test cell during a bremmstrahlung shot.
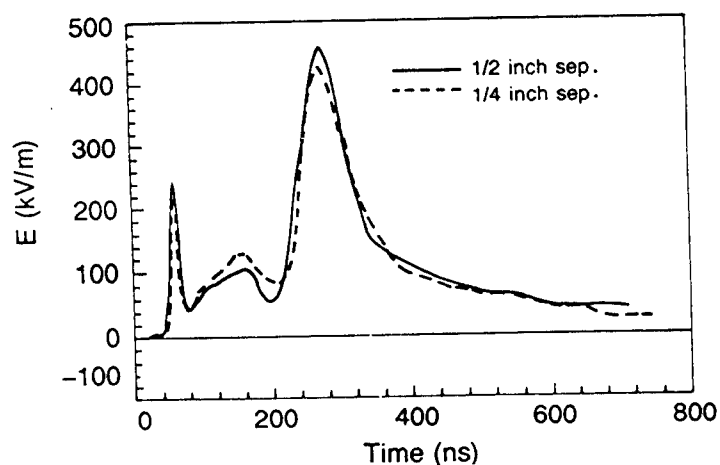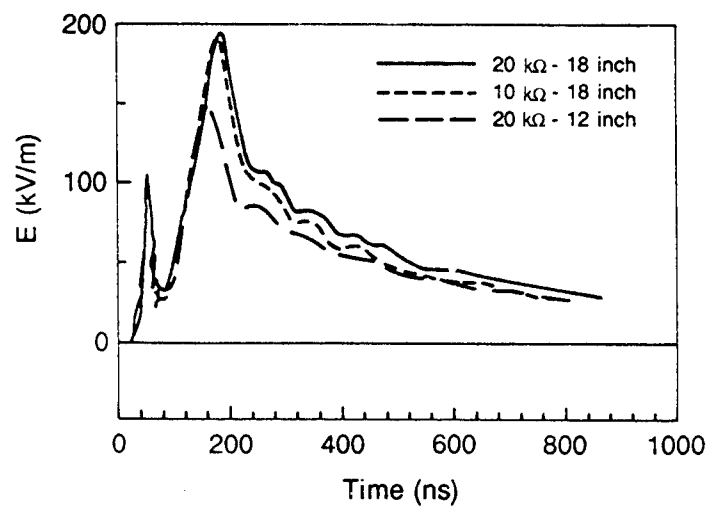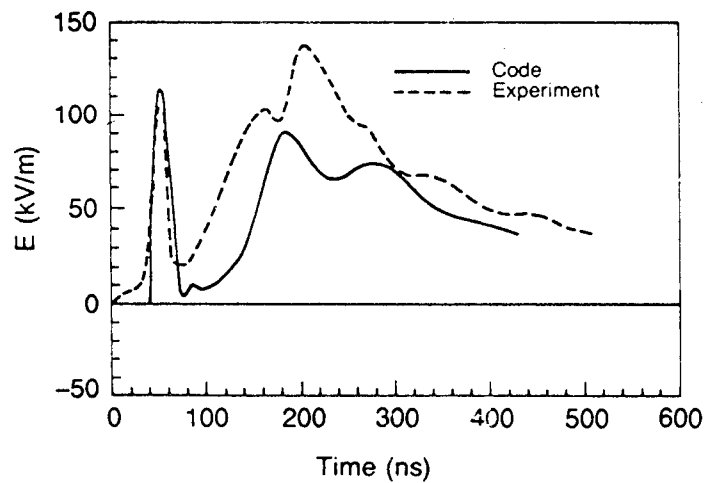

Figure 3. Electric field measurements made with wire grid sensors on the Aurora test cell floor using an e-beam that was not nose eroded or shortened by e-beam diversion.

230

Figure 4. Electric field measurements in the Aurora e-beam environment.

Figure 5. Comparison of finite difference computation and experimental measurement on Aurora test cell floor.

more TSR-like by shortening the E-field pulse and removing the valley.

To this end, the series of E-field measurements was made to ascertain whether or not the valley in the E-field waveform (Fig. 3 to 5) could be removed by shortening the e-beam pulse. Another goal was to determine whether the maximum E-field could be increased by decreasing the risetime. Drift tube risetime enhancement can be altered by changing gas pressure or gas species.

Discussion of Results:

Our discussion outlines how the e-beam current pulse risetime, duration, etc., can be manipulated to control the electric field produced in the test cell by the e-beam pulse. Figure 6 shows a series of E-field measurements obtained on the test cell floor by injecting a risetime enhanced diffuse e-beam into the test cell. The E-field measurements correspond to six different diverter settings. The range of E-field measurement varies from 20 to 500 kV/m. The range in E-field magnitudes can be explained in terms of the value of the injected electron beam current. As the diverter setting gets smaller, it not only decreases the pulse width, but also decreases the maximum current. We have, by shortening the injected e-beam pulses, removed the valley in the resulting electric field pulse, which was a major goal of the series of measurements shown in Figure 6.

Analysis of Aurora Test Cell Late-Time Ripple:

The long tails in the measured E-fields all show a sine wave ripple. This ripple is consistent with the fundamental cavity mode of the Aurora test cell. The Aurora test cell is, to a first approximation, a 5 x 12 x 20 m^3 rectangular cavity with a cylindrical perturbation: the drift tube. If we neglect the perturbation, the rectangular cavity mode with vertical E-field component responsible for the late time ripple has a wave number given by[15-17]

$$k_0^2 = \left(\frac{\pi}{12}\right)^2 + \left(\frac{\pi}{20}\right)^2$$

so the period is 80 ns. The result is in relatively good agreement with the experimental period of about 80 ns. The presence of other cavity modes could easily be investigated if we place E-field sensors on the side walls. Detailed magnetic field measurements could also be employed in a more precise and extensive cavity mode study.

The ripple is superimposed on a slowly decreasing late time electric field. As already indicated, the slowly decaying late time fields for the smaller diverter settings can be described in terms of a trapped or negative ionic charge. Electrons corresponding to a low acceleration voltage tend to have trajectories that terminate in the Aurora test cell and are responsible for the long after-pulse. It is possible to inject a relatively monoenergetic beam into the test cell by a combination of nose erosion and beam diversion. The energy of this beam can be controlled by adjustment of the Aurora Marx stage charge voltage and consequently the e-beam diode voltage. A high-energy monoenergetic beam produces a swarm of electrons that have trajectories with a low probability of termination in the volume of the test cell and subsequent capture by O_2. This results in a smaller late time electric field than would have been the case for a low energy beam. On the other hand, a low energy beam produces electrons with trajectories with a larger probability of termination in the test cell volume and consequently produces a relatively large slowly decaying field due to electron capture and ionic relaxation.
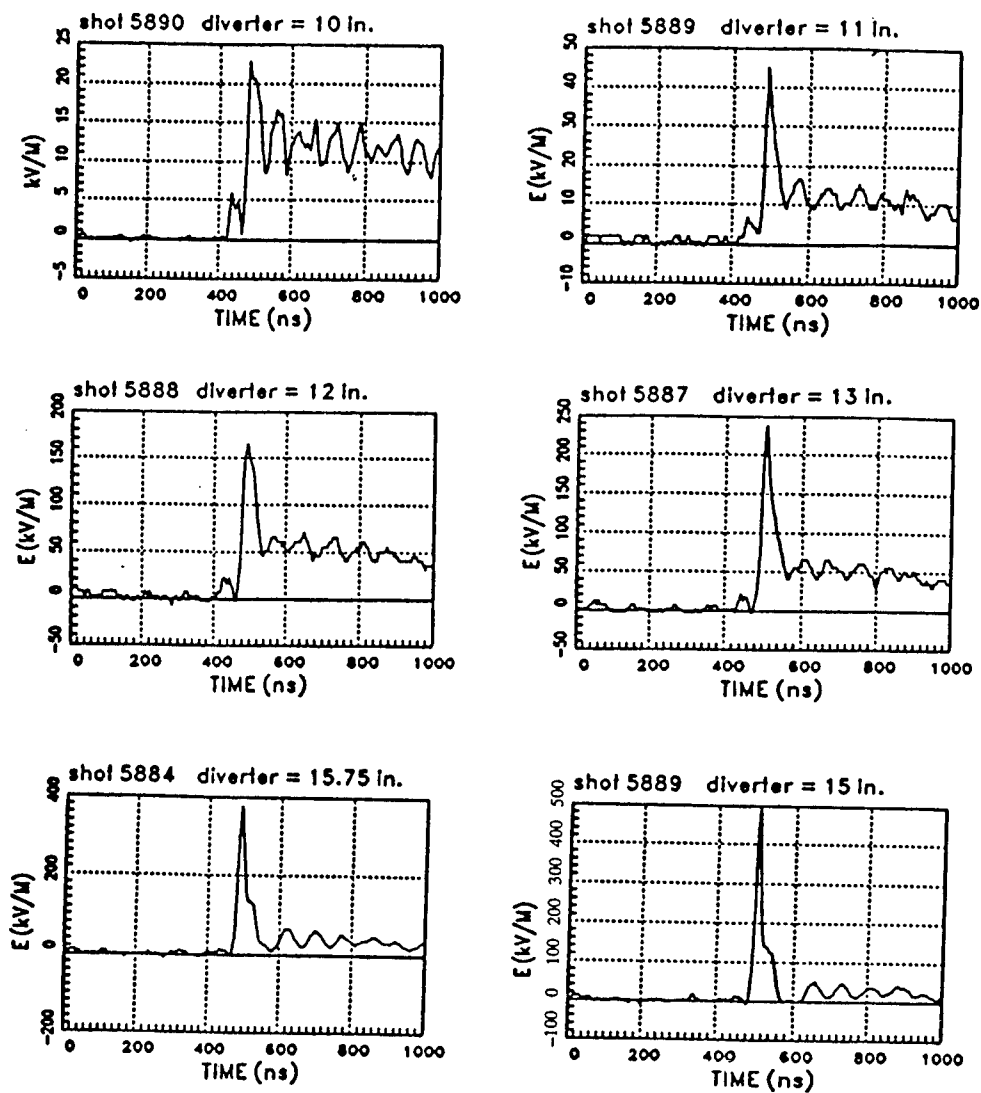
Figure 6. A series of electric field measurements obtained on the Aurora test cell floor using a nose-eroded diffuse e-beam.

Combined e-Beam and Bremsstrahlung Environments:

The Aurora test cell has been subjected to a number of combined e-beam and bremsstrahlung shots. Three of four conventional Aurora diodes were used to produce bremsstrahlung and one diode was employed to inject an e-beam directly into the test cell.

The three unmodified Aurora diodes produced relativistic electrons which then impinged on the unaltered Aurora thick tantalum bremsstrahlung targets. No risetime enhancement or beam diversion was employed in the bremsstrahlung diodes. Figure 7 shows an estimate of the bremsstrahlung distribution throughout the Aurora test cell produced with this three-target bremsstrahlung configuration. A number of electric field measurements were made throughout the Aurora test cell while the cell was subjected to the combined environment. The estimated bremsstrahlung dose rate is shown in Figure 7. We have obtained a TSR-like mixture of E-fields and dose rate.

Conclusions:

Much progress has been made in the simulation of the tactical source region. The Aurora pulse shape has been sharpened and shortened by using beam erosion and diverter techniques. A TSR-like environment has been produced in the Aurora test cell by combining e-beams with bremsstrahlung. In fact, the Aurora is a superb test bed that can be employed to develop concepts that will be applied in the development of the Army's future TSR simulator. A pure bremsstrahlung ionizing radiation source, however, would be preferable to an e-beam source. Presently available bremsstrahlung sources do not have the necessary magnitude over as large a volume as an e-beam source or mixed e-beam/bremsstrahlung source. Future work will involve removal of the jitter between the four Aurora diodes, the incorporation of risetime enhancement in each of the diodes, and the reorientation of the diodes. With these refinements, the bremsstrahlung spatial isodose-rate distribution (shown in Fig. 7) can be made more homogeneous. The bremsstrahlung dose-rate pulse shape will then also have a controllable risetime and falltime.

REFERENCES

1. Conrad L. Longmire, "On the Electromagnetic Pulse Produced by Nuclear Explosions," IEEE Trans. on Antennas and Propagation, AP-26, No. 1, January 1978.

2. M. Bushell, R. Manriquez, G. Merkel and W.D. Scharf, "Aurora Flash X-Ray Facility as a Source-Region EMP Simulator," Army Science Conference Proceedings, 15-18 June, 1982, p. 159.

3. B. Bernstein and I. Smith, "AURORA, An Electron Accelerator," *IEEE Trans. Nucl. Sci.*, NS-20, No. 3 June 1973.

4. M. Bushell, J. Deppe, K. Kerris, G. Merkel and D. Whittaker; and L.E. Stevenson and R.B. Miller, "Hermes II SREMP Simulation Experiment," *Journal of Radiation Effects (HEART)*, Vol. 3, No. 1, Winter 1984, published June 1985, p. 97.

5. Conrad Longmire, "General Considerations on Use of AURORA for Simulating EMP Coupling," AURORA EMP Memo 1, MRC, Santa Barbara, CA, Sept. 22, 1975.
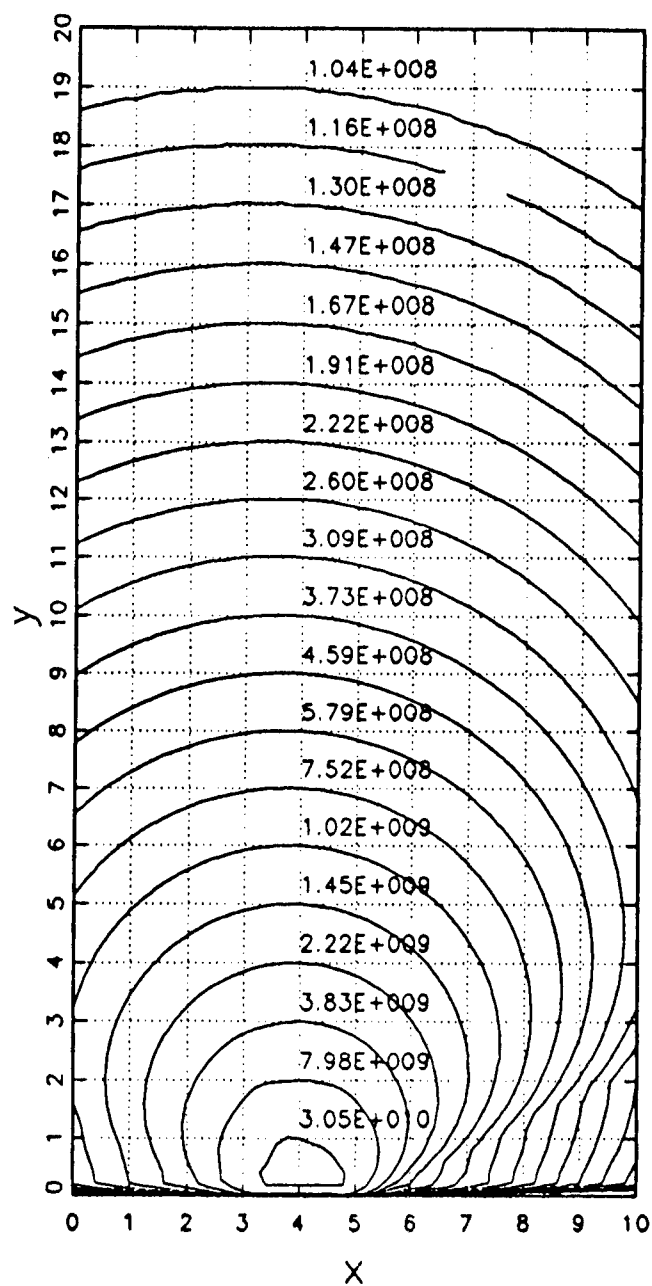
Figure 7. Bremmstrahlung dose rate produced by three of the Aurora bremmstrahlung targets.

6. S.E. Graybill, "AURORA Electron Beam Modifications," HDL-TR-1862, July 1978.

7. G. Merkel, W.D. Scharf, and D. Spohn, "Source Region EMP Simulator-A Parallel Plate Transmission Line in the AURORA Test Cell," DNA Symposium on EMP Simulation and System Hardening, Monterey, CA, October 28-29, 1980.

8. M. Bushell, S.E. Graybill, K.G. Kerris, G. Merkel, W.D. Scharf and D.A. Whittaker, "The Direct Injection of Electron Pulses into Air-A SREMP Simulation Tool," Third IEEE International Pulsed Power Conference, June 1981, Albuquerque, NM (appears in *Digest of Technical Papers*, p. 513, 1981).

9. M.B. Bushell, W.O. Coburn, G. Merkel and W.D. Scharf, "Experimental and Theoretical Evaluation of Electromagnetic Environment Produced by Direct Electron Injection at AURORA," Fourth IEEE International Pulsed Power Conference, June 6-8, 1983, Albuquerque, NM (appears in *Digest of Technical Papers*, p. 41).

10. George Merkel and William D. Scharf, "Source Region Electromagnetic Pulse Simulator," U.S. Patent #4,393,509, July 12, 1983.

11. M. Bushell, R. Fisher, D.W. Merewether, G. Merkel and W.D. Scharf, "Predicted and Measured Response of a Missile Model in the AURORA e-Beam Environment," *IEEE Trans. Nucl. Sci.*, NS-30, No. 6, p. 4552, December 1983.

12. M. Bushell, R. Manriquez, G. Merkel and W.D. Scharf, "AURORA Test Cell Electron Beam Environment Response of a Large Loop," *IEEE Trans. Nucl. Sci.*, NS-30, No. 6, p. 4558, December 1983.

13. M. Bushell, S. Graybill, M. Litz and G. Merkel, "Enhanced Risetime Bremsstrahlung AURORA Environment," *IEEE Trans. Nucl. Sci.*, NS-31, p. 1299, December 1984.

14. M. Bushell, S. Graybill, M. Litz, G. Merkel and H. Roberts, "Excitation of a Canonical RF Shelter with Relativistic Electrons with Enhanced Risetime Bremsstrahlung," *IEEE Trans. Nucl. Sci.*, NS-32, p. 4320, December 1985.

15. R. Fleetwood, G. Merkel, H. Roberts, M. Smith, M. Bushell and K. Kerris, "IEMP Excitation of an S-250 RF Shelter: e-Beam versus Bremsstrahlung," *Journal of Radiation Effects (HEART)*, Vol. 5, No. 1, p. 136, 1987.

16. R. Fleetwood, K. Kerris, G. Merkel, H. Roberts and M. Smith, "Electromagnetic Excitation of a Generic Cavity with a Variable e-Beam Pulse," *IEEE Trans. Nucl. Sci.*, NS-34, p. 1521, December 1987.

17. R. Fleetwood, K. Kerris, G. Merkel, H. Roberts, and M. Smith, "Ionizing Radiation Dose Distribution Produced Inside a Canonical Cavity by Direct e-Beam Excitation, *IEEE Trans. Nucl. Sci.*, NS-34, p. 1521, December 1987.

Ground Water Resource Assessments

Dwain K. Butler, Ph.D.
US Army Engineer Waterways Experiment Station
Vicksburg, Mississippi 39180-6199

Background

Potable water is a prime requisite for maintenance of life. In many arid and semiarid regions, adequate water supply has long been difficult to maintain. In many agricultural and industrial settings, maintaining an adequate water supply is becoming a critical problem, even in temperate and tropical regions. As surface water supplies become more and more taxed to meet demands and surface water contamination decreases the available sources of potable water, man is forced to turn to exploitation of ground water reserves at an ever-increasing rate. In 1980, ground water supplied 35 percent of drinking water, 80 percent of rural domestic water, 40 percent of irrigation water, and 6 percent of self-supplied industrial water[1] There is great demand for optimization of ground water utilization from existing well fields and protection and replenishment of the resource. The increasing water demand is a driver for location and development of new ground water resources. Ground water exploration programs can aid well-field design and well placement.

Adequate water supply is a critical requirement for support of military field operations and for fixed military bases. For field operations, rapid development of ground water resources may be desirable in regions where existing water supplies are insufficient or unreliable, particularly in arid and semiarid regions. Standard procedures for ground water exploration are not directly suited for rapid ground water detection and development. There is no single device, now and for the foreseeable future, which can be utilized to determine the depth to potable ground water at a given location with a high confidence level. Dreams of such a "black-box" device lie more in the realm of "water witching or divining" than sound geoscience technology. Ground water is usually "detected" as a matter of course in field ground water exploration investigations that are designed to map the complete hydrogeologic regime. In 1981, a Water Supply Task Force of the Defense Science Board identified ground water detection as a major technology shortfall, and the Joint Chiefs of Staff established ground water detection capability as a high-priority military requirement.

A Ground Water Detection Workshop held at WES in 1982[2] concluded
(1) that there are two currently fieldable geophysical methods,
electrical resistivity and seismic refraction, which offer a near-term
solution to the required ground water detection capability, and (2) that
there are several state-of-the-art and emerging geophysical techniques
that have intermediate- to far-term potential for application to ground
water detection. The Army's Water Detection Response Team (WDRT) was
formed in 1985 to provide rapid response ground water location capa-
bility. The WDRT consists of a Database Element, a Remote Sensing
Element, a Geophysics Element, and a Supporting Specialists Element. The
Geophysics Element maintains the capability to field the near-term
detection capability and monitors emerging technology for possible future
WDRT utilization.

Recently, the problem of contamination of ground water by military
operations and hazardous and toxic waste disposal has emerged as part of
an overall, high-priority national environmental concern. Programs for
environmental protection and installation restoration/remediation have
emerged in response to this high-priority problem. These programs are
guided by the desire to protect public health and welfare, to protect the
environment for future generations, to maintain continued, safe, long-
term use of military installations, and to provide for safe turnover of
military lands to the public sector through base closure. These programs
require an assessment of the ground water resource for the presence of
contaminants. The assessments utilize geophysics, novel penetrometer
sensor systems, ground water sampling/monitoring wells, and ground water
modeling.

Ground water detection, exploration, and assessment are the thrusts
of several past and current research efforts at the US Army Engineer
Waterways Experiment Station (WES). The WES also performs field
investigations in support of military water supply requirements,
installation restoration, and base closure. Sponsors for the WES
research efforts and field investigations are varied, both military and
nonmilitary. The WDRT's Geophysics Element and Supporting Specialists
Element are located at the WES. This paper concentrates on the major
contributions of geophysics to ground water resource assessments.

Hydrogeology

Exploration for ground water involves either direct or indirect
methods, with the goal of defining or characterizing the hydrogeology of
the area of interest. Direct methods include surface geological recon-
naissance and drilling. The act of drilling can be considered sub-
surface geological reconnaissance; also, drilling allows the application
of borehole geophysical logging for subsurface characterization.
Drilling is expensive and time consuming, and alone is not an efficient

method of exploration. Indirect methods, properly classified as remote sensing methods, involve surveys conducted on the earth's surface or from airborne or satellite (aerial) platforms. The aerial remote sensing methods generally attempt to map or detect structural, geomorphic, and vegetative <u>surface</u> indicators of the presence of ground water. Some radars and other electromagnetic geophysical methods can actually achieve limited penetration below the earth's surface from aerial platforms and thus provide subsurface indicators of the presence of ground water. Surface remote sensing or geophysical methods attempt to detect and map structural, stratigraphic, and aquifer property <u>subsurface</u> indicators of the presence of ground water.

When usable quantities of ground water are present in a geological material, it is called an aquifer. The key intrinsic properties of an aquifer are its porosity and permeability, which describe its capability to store and transmit water, respectively. Other important properties of an aquifer are its thickness, areal extent, and areal continuity, which determine the total quantity of ground water held in storage. Various types of aquifers (hydrogeological models) are illustrated in Figures 1-4. Surface geological reconnaissance can detect areas of ground water recharge and discharge but, in the absence of borehole information, can only infer ground water occurrence. Aerial remote sensing can infer areas of ground water recharge and discharge and areas with very shallow ground water from vegetation patterns. Also, aerial remote sensing methods are useful for detecting lineaments which may represent surface expressions of fractures (see Figure 3); intersections of fractures are desirable locations for wells. Surface geophysics is used to verify aerially mapped lineaments/fractures and provide precise well positioning in the fracture zone. Mapping of the details of the hydrogeological models in Figures 1, 2, and 4 requires surface geophysics and drilling. Figure 4 illustrates the fact that knowledge of the hydrogeology is critical for location of water wells; drilling depth, water production rate, and useful well lifetime are highly location dependent. Drilling sites to the far left of the model in Figure 4, for example, will result in a dry hole due to low permeability and ground water storage capacity of the rock.

Ground Water Exploration

Geophysical methods are routinely used throughout the world in exploration programs for the assessment and development of ground water resources. The geophysical methods that are predominantly used in these programs are gravity, electrical resistivity, electromagnetic, and seismic refraction methods. Occasionally only one of these methods will be used in exploration, but generally at least two of the methods are used in a complementary program. A geophysical groundwater exploration program will use all available borehole and other geological data to
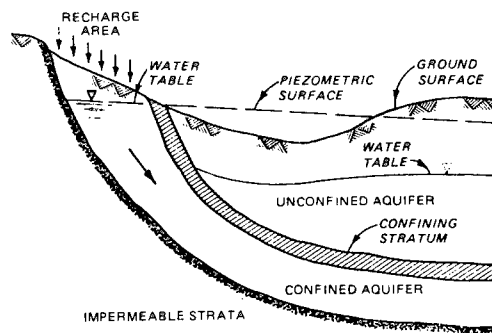
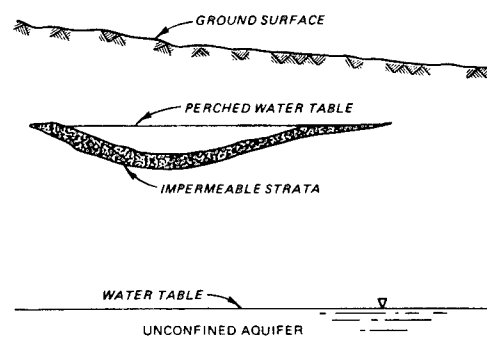Figure 1. General model for ground-water recharge and occurrence.
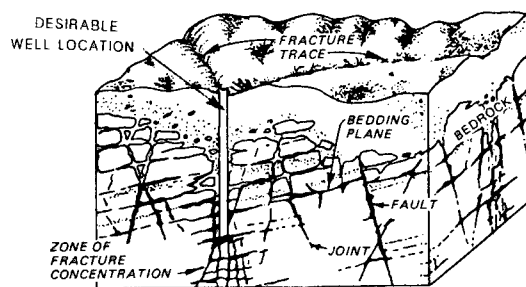

Figure 2. Perched ground-water.

Figure 3. Fractured rock aquifer.

① WELL LOCATION TO TAP GROUND WATER IN FRACTURED ROCK
② WELL LOCATION TO TAP PERCHED GROUND WATER
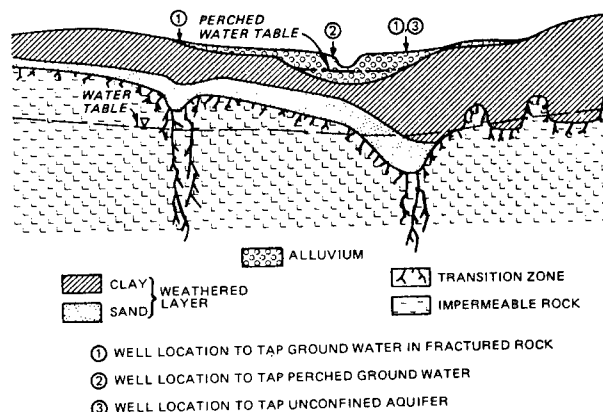③ WELL LOCATION TO TAP UNCONFINED AQUIFER

Figure 4. Composite ground-water occurrence model and possible well sites.
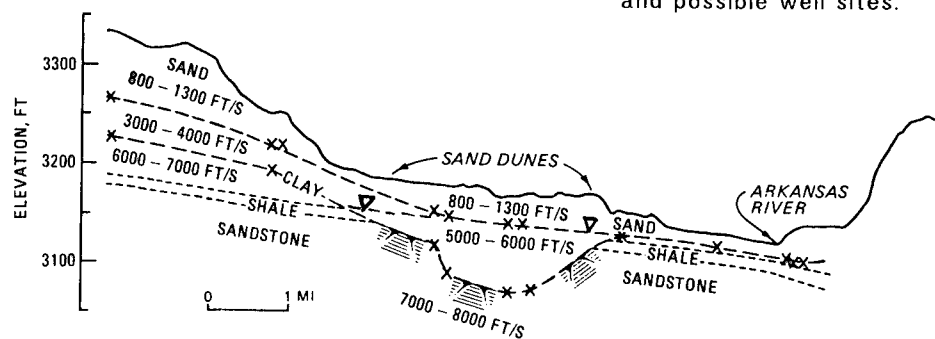

Figure 5. Geologic cross section resulting from seismic refraction surveying and selective drilling.

240

produce the best possible assessment of the ground water potential in an area. Figure 5 illustrates the results of a seismic refraction survey in Western Kansas which delineates a buried channel.[3] The portrayal in Figure 5 is a geophysical model which has been interpreted to give a geological/hydrogeological model of the subsurface beneath the survey line. In this case, the water table was actually detected in the central part of the survey profile by the occurrence of a seismic compression wave velocity that is characteristic of a saturated, uncon- solidated sedimentary material (1,500 m/s or 5,000 ft/s). Even in the absence of the aquifer property indicator, the stratigraphic/structural indicators suggest the greatest ground water potential in the center of the buried channel. Identification of material types was made by correlation with exploratory borings near each end of the survey line.

Results of a ground water exploration program at Osan Air Force Base (AFB), Korea, consisting of closely spaced seismic refraction and electrical resistivity surveys are shown in Figure 6.[4] The problem at Osan AFB was that some wells would stop producing useful quantities of water after a very short period, while other wells produce for years. The geophysical model reveals that the shallow aquifer at Osan AFB is not laterally continuous but is "interrupted" by clay "plugs," and wells located in some of the smaller "pockets of ground water" in the clay will quickly cease to produce water. The complementary electrical resistivity and seismic compressional wave velocity for a given region
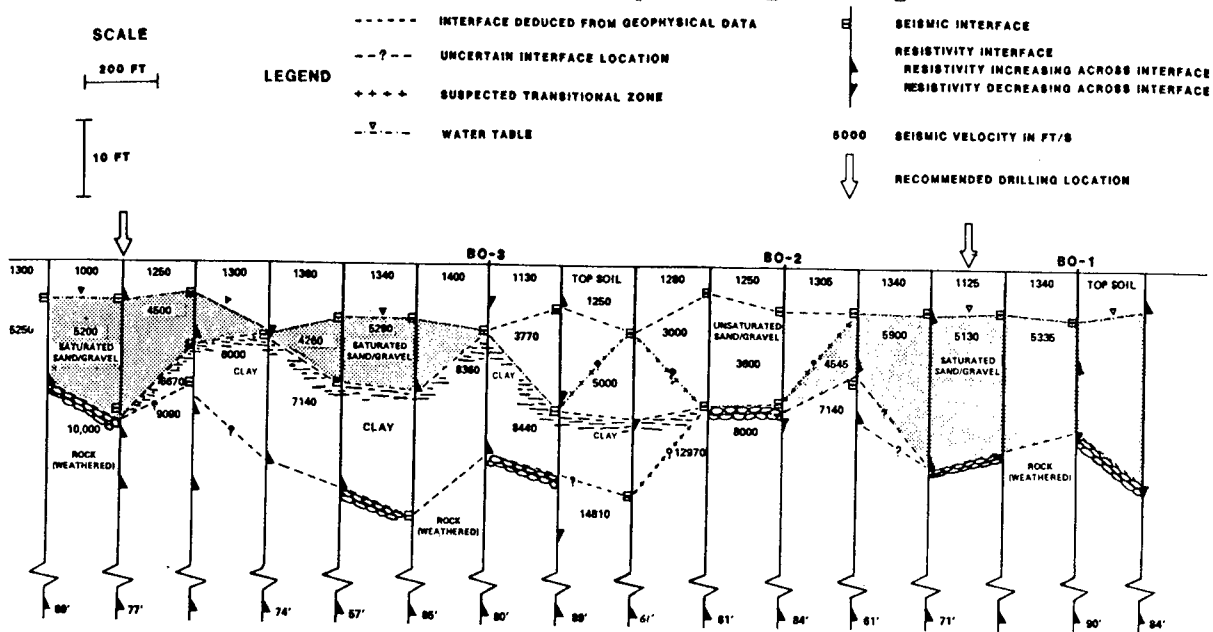

Figure 6. Geophysical cross-section, ground-water supply study, Osan Air Force Base, Korea.

of the cross section allows generic material type and hydrogeological property identification, while correlation with selected exploratory boring results allows specific material type identification. Recommended well locations are indicated in Figure 6, based on the geophysical, stratigraphic and aquifer property indicators.

Ground Water Detection

The expression "ground water detection," in contrast to ground water exploration, applies to the concept of actually detecting/interpreting the presence (or absence) of ground water and the depth to the water table (or confined aquifer) beneath a given location on the surface by conducting one or more types of geophysical tests at that location. In the ideal case, the aquifer thickness and water quality would also be determined. For some cases, information on ground water occurrence and other geological factors might be available; but, for the general detection scenario, the assessment of the presence of ground water must rely solely on the geophysical results. For many anticipated deployments of the WDRT, for example, it is envisioned that the geophysical surveys would be conducted to aid in choosing between alternate possible drilling sites in an area already identified as having ground water potential by other methods.

The following tabulation presents a "shopping list" of geophysical methods and their applicability to ground water detection and exploration:

Geophysical Method	Detection	Exploration/Assessment
Seismic Refraction	X	X
Seismic Reflection		X
Seismic V_p/V_s [*]	X	
Electrical Resistivity	X	X
Transient Electromagnetic	X	X
Ground Penetrating Radar	?	X
Other Electromagnetic	?	X
Magnetic		X
Gravity		X
Aerial (Gravity, Magnetic, EM)		X

[*] Compression wave velocity -- V_p; shear wave velocity -- V_s.

Seismic Refraction

The seismic refraction method uses a surface energy source that generates a seismic disturbance which propagates into the subsurface.

This disturbance undergoes reflections and refractions at interfaces between different materials. The seismic refraction method utilizes seismic waves which refract at the critical angle and propagate along the interfaces. Data from the refraction survey are interpreted to give depths to interfaces and seismic velocities of the layers between the interfaces (strata). The physical principle involved in the detection of the water method by the seismic refraction method is that V_p will increase from 300-700 m/s to 1,375-1,675 m/s at the water table (compression wave velocity contrasts of 3-5 typical), where the water table occurs at shallow depths (< 50 m) in unconsolidated sediments (silts, sands, and gravels). In fact the occurrence of a characteristic 1,500 m/s at a location is generally strongly indicative of a ground water table, although some weathered rocks and massive clays can have this velocity also. Occurrence of this characteristic compression wave velocity at the water table is illustrated in Figure 5. If the water table occurs at greater depths, the seismic velocity of the saturated sediments can be as high as 2,300-2,500 m/s; but in these cases, the velocity of the unsaturated sediments above the water table can be as high as 1,200 m/s (velocity contrast of 2 typical). The smallest compression wave velocity contrast will occur when the water table occurs in very fine-grain sediments. When a water table occurs in rock, there will be a velocity increase across the interface, but the contrast may be small. For a confined rock aquifer, there may be little in the seismic compression velocities determined from a standard refraction survey to suggest the presence of ground water without independent or complementary information.

Conceptually, the compression wave velocity increase at the water table can be explained by the following expression derived from elastic wave propagation theory:

$$V_p = ((K + 4G/3)/d)^{1/2}, \qquad (1)$$

where K is the bulk modulus, G is the shear modulus, and d is the bulk density. Addition of water to an unconsolidated sediment results in an increase in K by an order of magnitude or more, with only a 15-20 percent increase in d and virtually no change in G. The result of the changes in K and d is the dramatic increase in V_p at the water table. The shear wave velocity is given by:

$$V_s = (G/d)^{1/2}. \qquad (2)$$

It follows that the shear wave velocity does not change significantly (may actually decrease) at the water table. Both K, G, and d will change at an interface between two different rock types in such a way that the ratio of V_p/V_s in the materials on each side of the interface may change

very little. However, it is clear that V_p/V_s will change dramatically at the water table and is independent of d. Thus V_p/V_s will be highly indicative of the occurrence of ground water and will increase the differentiability of the water table from weathered rocks and clays. In principle, it is possible to determine both V_p and V_s from seismic refraction as well as seismic reflection surveys; field procedures and interpretation methods are active research areas.[5,6]

Electrical Resistivity

The electrical resistivity method, which is applicable to ground-water detection is vertical resistivity sounding; where the objective is to make electrical measurements at the surface from which the vertical variation of electrical resistivity with depth can be interpreted. Resistivity is a fundamental physical property of geological materials. The range of resistivities for geological materials is 10^{20} ohm-m, although the range commonly encountered in ground water exploration and detection is 10^5 ohm-m. This large range for resistivity is in contrast to the variation by a factor of approximately 15 observed in V_p.

Most soils and rocks conduct current primarily electrolytically, i.e., through interstitial pore fluid. Thus porosity, water content, and dissolved electrolytes in the water are the dominant factors in controlling resistivity rather than the soil or rock type. Clays, which can conduct current both electrolytically and electronically, are a major exception to this generalization. The general relation between the bulk resistivity RHO_b and the porosity n (volume fraction), pore fluid saturation S_w (volume fraction of n), and pore fluid resistivity RHO_w can be expressed by the empirical relation known as Archie's Law:

$$RHO_b = a \times RHO_w \times n^{-m} \times S_w^{-p}, \tag{3}$$

where a, m and p are constants which depend on the soil or rock type but vary within relatively small ranges. Below the water table $S_w = 1$ (100 percent saturation). It is $RHO_b(z)$, where z is depth below the surface, that is interpreted from an electrical resistivity sounding. Since representative values of a and m can be selected, the aquifer porosity can be calculated from values of RHO_b if values of RHO_w are known or can be estimated. Alternatively, if n is known or can be estimated, then RHO_w can be calculated. The porosity n is a fundamental hydrogeological parameter, while RHO_w can be interpreted in terms of water quality. A common use of resistivity sounding is for detecting the fresh water/ brackish (salt) water interface, which will be indicated by the occurrence of a prominent resistivity decrease. The water table will be detected as the top of a conductive or less resistive layer; since, except for unusual conditions, even fresh potable ground water is lower in resistivity than the dry aquifer material. A well was successfully sited

by the WDRT during Bright Star 85 in Egypt using electrical resistivity.

Complementary Methods

The electrical resistivity and seismic refraction methods are complementary in the sense that they respond to or detect different physical properties of geologic materials.[5,7] Both methods can detect the water table under certain conditions, and in these cases one method serves to confirm the results of the other or to resolve ambiguities. Certain conditions, however, such as the presence of a fresh water/brackish water interface, can be detected by one method but not the other. For the common setting of an alluvial, unconfined (water table) aquifer overlying a relatively impermeable basement rock, the combination of seismic refraction and electrical resistivity surveys can often yield a nearly complete hydrogeological interpretation at a location: depth to water table; depth to brackish water, if present; depth to impermeable basement; porosity estimates. For other geologic settings, the qualitative values of V_p and RHO_b, in addition to characteristic values and ratios at interfaces, give considerable insight to possible soil/rock type and hydrogeological properties as indicated in the following tabulation:

V_p	RHO_b	Qualitative Interpretation
High	High	Impermeable rock.
High	Interm.	Rock. Possible aquifer.
High	Low	Rock. Possible aquifer; probably brackish.
Interm.	High	Dry, unconsolidated sediments at depth; dry, weathered, or fractured rock;
Interm.	Interm.	Possible aquifer in uncons. sediments or in weathered rock.
Interm.	Low	Clay or brackish water.
Low	High	Dry unconsolidated sediments; no clay.
Low	Interm.	Clayey, uncons. sediments; wet sediments.
Low	Low	Wet, clayey sediments.

V_p -- High (>3,000 m/s); Low (<1,000 m/s)
RHO_b -- High (>300 ohm-m); Low (<10 Ohm-m)

Successful applications of complementary methods for ground water detection have been accomplished at White Sands, NM, at Fort Carson, CO, by the WDRT during Gallant Eagle 86, Fort Irwin, CA, and by the WDRT during Bright Star 87 in Somalia. The cross section in Figure 6 was deduced by a complementary methods approach.

Electromagnetic Methods

If a device is ever developed that approaches the ideal of a

black-box" water detector, it will likely be an electromagnetic (EM) device. There are numerous EM geophysical techniques, ranging from near-DC induction techniques to GHz wave propagation techniques. It is possible that some innate property of an aquifer/ground water system will ultimately be amenable to interrogation or probing by an EM technique and allow direct ground water detection. However, direct ground water detection must be viewed as a long-term goal. The immediate application of the EM techniques is as a replacement or supplement to the electrical resistivity technique.

The electrical resistivity techniques require ground contact, generally by driving metal rods into the ground. This procedure is labor intensive and in some settings, particularly arid and semiarid areas, it is difficult to achieve good results due to high rod to ground contact resistance. The EM techniques do not require ground contact and are logistically less cumbersome and labor intensive than electrical resistivity techniques. Like the electrical resistivity techniques, EM surveys are interpreted to give resistivity or conductivity (= $1/RHO_b$) of subsurface materials. Continuous wave EM (CWEM) techniques are useful for rapid horizontal profiling, to assess horizontal variations between electrical resistivity sounding or drillhole locations.[8] Transient EM (TEM) techniques are useful for vertical sounding, and therefore could replace electrical resistivity sounding in ground water exploration or detection programs.[9]

CWEM techniques use a low frequency (typically 1-20 KHz) transmitter coil to induce secondary currents and magnetic fields in subsurface conductors. The secondary magnetic fields are detected by a receiver coil as a percentage of the transmitter primary magnetic field. The apparent conductivity of the subsurface is determined as a function of the transmitter frequency, the transmitter-receiver coil spacing, the coil orientations, and the measured secondary to primary magnetic field ratio. Different depths of investigation can be achieved with the CWEM technique by varying the transmitter frequency and the transmitter-receiver coil spacing. In general, depth of investigation increases as transmitter frequency decreases and as coil spacing increases. The CWEM method was successfully demonstrated for ground water exploration and detection at White Sands, NM, and at Fort Carson, CO.[8]

A typical TEM sounding configuration utilizes a small receiver coil centered within a larger transmitter coil as illustrated in Figure 7. The transmitter coil will typically be square with side lengths of 40-100 m. A sequence of current pulses is injected into the transmitter, with alternating polarity and rapid turnoff. The current pulse generates a primary magnetic field which induces currents in the subsurface. The induced currents resemble a smoke ring which diffuses downward and outward in the subsurface. The diffusing current ring generates secondary
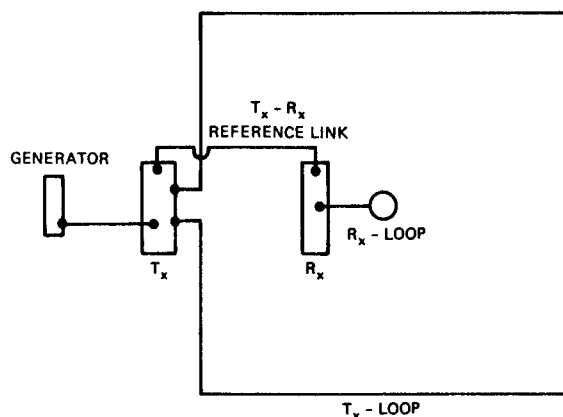
Figure 7. Simplified, schematic illustration of a TEM system.

magnetic fields which are dependent on the electromagnetic properties, including the conductivity, of the geologic materials. The motion of the current ring as a function of time is portrayed in Figure 8 for various subsurface conductivities.[9] The receiver loop detects and samples the secondary magnetic fields as a function of time while the primary field is turned off. Figure 9 illustrates the transmitter pulse sequence, electromagnetic induction by the transmitter current pulse turnoff, and measurements of the secondary magnetic field during transmitter off-time by a receiver which takes 20 measurements (samples or gates) of the inductive transient. Depth of investigation with the TEM method increases as the transmitter magnetic moment increases and as the measurement time of the transient increases (due to the downward diffusing of the current ring and hence the zone from which the secondary magnetic fields originate). That is, early time measurements correspond to shallow depth properties and later time measurements to properties from successively greater depths. The TEM method has been successfully applied to ground water detection exercises at White Sands and at Fort Iwrin. Reference 9 contains an extensive comparison of TEM, CWEM, and electrical resistivity results with known hydrogeology at several locations at White Sands. Based on requirements for ground water detection and exploration and experience gained from field exercises, new instrumentation has been developed to allow optimized application of the TEM method in the important 5 m to 150 m depth range, and interpretation programs have been developed which can be used on microcomputers.

Ground Water Contamination

Application of geophysics to detection and assessment of ground-water contamination can be illustrated by the results of an investigation at Radford Army Ammunition Plant (RAAP), Radford, VA.[10] The site
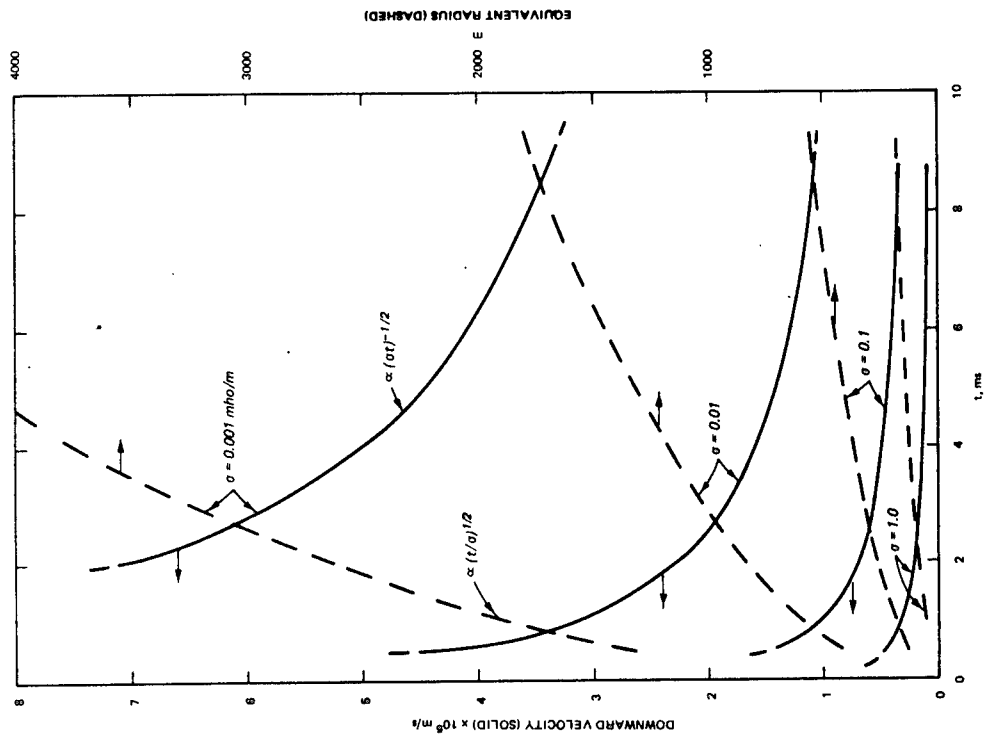
a. TRANSMITTER CURRENT WAVEFORM

b. GATE LOCATION AND WIDTHS (30 AND 3Hz)

c. CURRENT IN TRANSMITTER LOOP

d. INDUCED ELECTROMOTIVE FORCE CAUSED BY T CURRENT

e. SECONDARY MAGNETIC FIELD CAUSED BY EDDY CURRENTS

Figure 9. Concepts of the transient electromagnetic method.

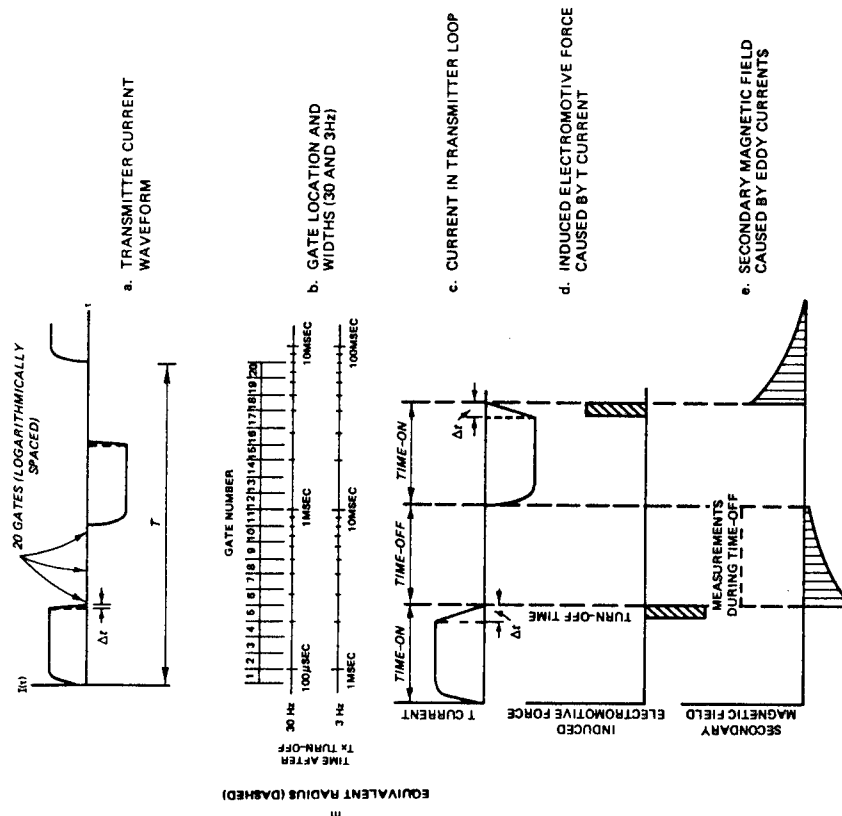Figure 8. Motion of equivalent current ring induced in a half-space of conductivity σ by a loop transmitter.

248

investigated at RAAP contains several hazardous waste burial trenches.
Exact location and number of trenches was not accurately known. The
purposes of the geophysical survey program were to locate and map the
trenches, map possible sinkholes near or beneath the site, and detect and
map any contaminant plumes leaving the site. The site is underlain by
limestone in which karstic features (e.g., sinkholes and cavities) are
common. Figure 10 shows suspected locations of burial trenches and the
layout of CWEM survey lines; lines EM-1 through EM-11 were located to
detect and map some of the trenches, while line EM-12 was located to
detect possible contaminant plumes leaving the site. Figure 11 shows the
EM signatures of the trenches and contents for one EM coil spacing,
revealing trench locations and conductivity variations within the
trenches. Finally, Figure 12 shows depth contours of the top of rock
determined by seismic refraction surveys, which reveal the signature of a
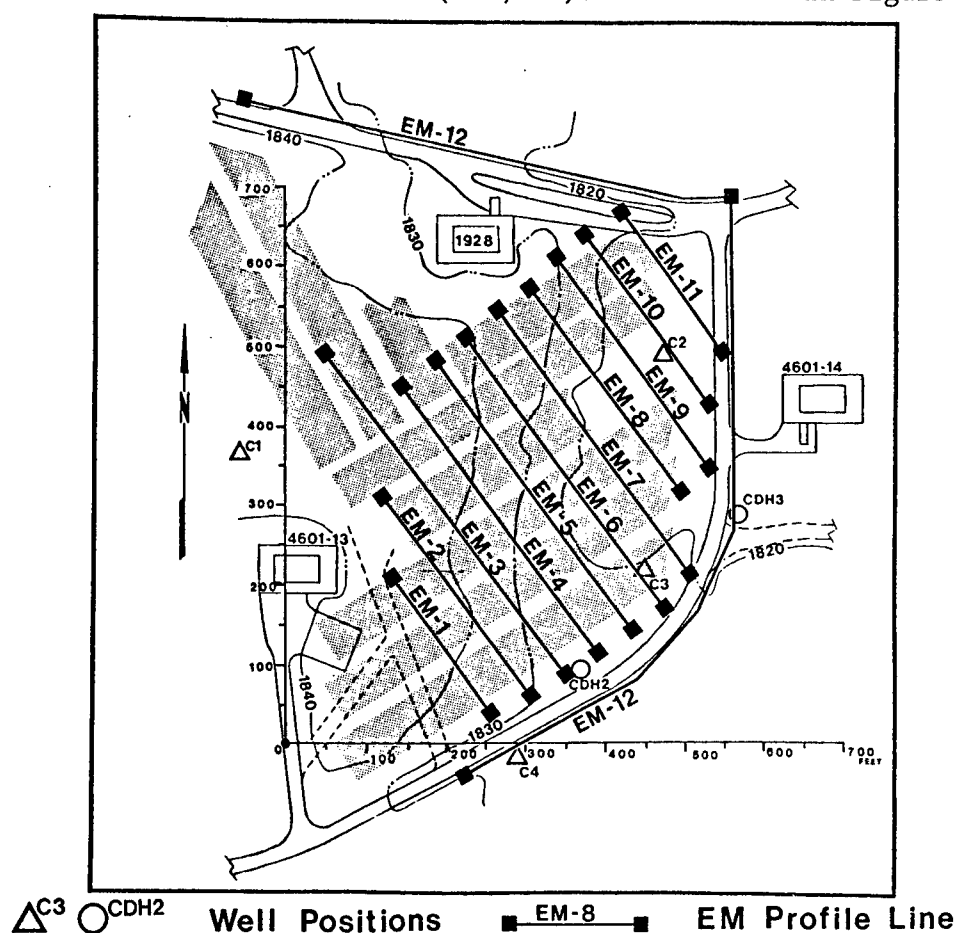sinkhole centered at location (500,350). Also shown in Figure 12 is the

Figure 10. Suspected burial trench locations and EM survey lines.
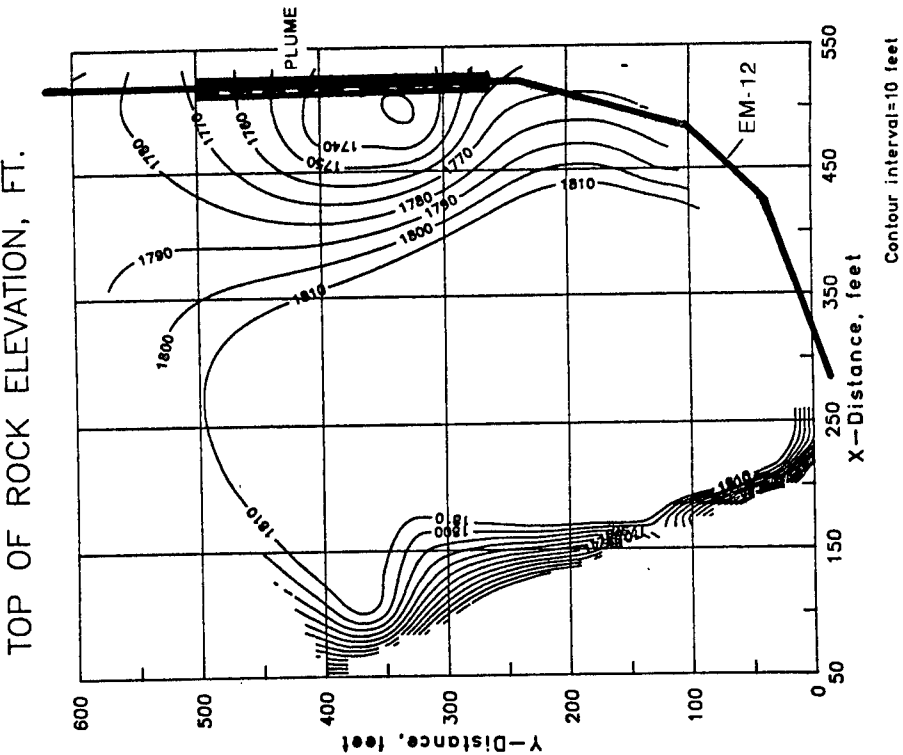
TOP OF ROCK ELEVATION, FT.

Figure 12. Top of rock elevation contours; EM-12 survey line shown with location of interpreted contaminant plume indicated.
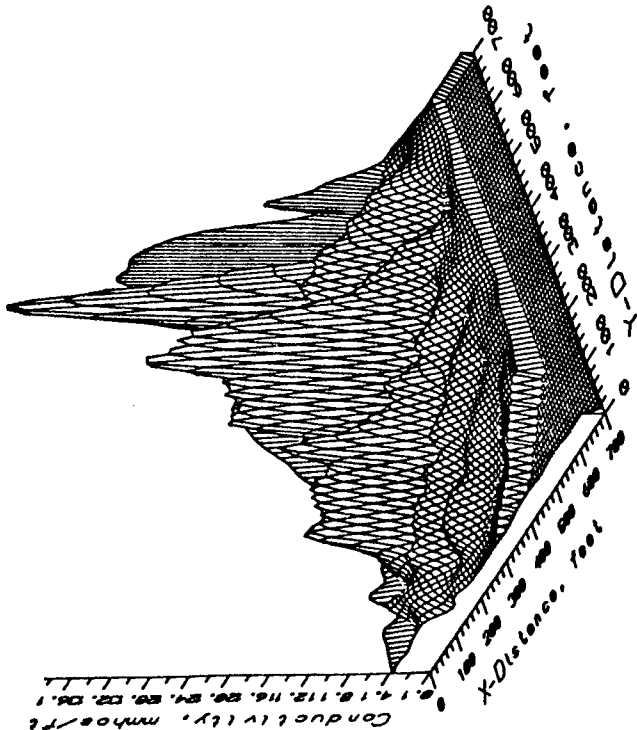
Figure 11. Perspective plot of EM conductivity survey results showing trench signatures; Radford Army Ammunition Plant.

location of a contaminant plume interpreted from the results of the EM-12 survey. The coincidence of the plume location with the sinkhole raises the possibility that the contaminants are flowing into the sinkhole and into the limestone formation beneath the site. The geophysical survey results can be used to plan selected, representative drilling and sampling of trench contents and for optimal placement of monitoring wells for off-site contaminant migration detection.

Conclusions

Through experience gained from basic research programs, the Army's Water Detection Response Team, and reimbursible work for a variety of sponsors, a general capability has been developed for assessment of ground water resources. Assessments include general ground water exploration, specialized ground water detection requirements, and ground water contamination detection and mapping.

References

1. AIPG, 1984. "Ground Water, Issues and Answers," Amer Inst of Prof Geol, Arvada, CO.
2. Dardeau, E.A. 1984. Proceedings of the Ground Water Detection Workshop, 12-14 January 1982, Vicksburg, Mississippi, US Army Engineer Waterways Experiment Station (CEWES), Vicksburg, MS.
3. Butler, D.K., and Bruhn, R.W. 1971. "Seismic Refraction Survey of a Proposed Dam Site near Kendall, Kansas," MRDL Letter Rpt, US Army Engineer Division Laboratory, Missouri River, Omaha, NE.
4. Butler, D.K., Yule, D.E., and Kean, T.B. 1981. "Osan Air Force Base, Korea, Water Supply Study," Misc Paper GL-81-9, CEWES.
5. Butler, D.K., and Llopis, J.L. 1985. "Military Requirements for Geophysical Ground Water Detection and Exploration," Proc Conf on Surface and Borehole Geophysical Methods in Ground Water Investigations, National Water Well Association.
6. Schuyler-Rossie, C. 1987. "Military Hydrology, Report 15, The Seismic Refraction Compression-Shear Wave Velocity Ratio as an Indicator of Shallow Water Tables, A Field Test," Misc Paper EL-79-6, CEWES.
7. Butler, D.K., and Llopis, J.L. 1984. "Military Hydrology, Report 6, Assessment of Two Currently Fieldable Geophysical Methods for Ground Water Detection," Misc Paper EL-79-6, CEWES.
8. Butler, D.K. 1986. "Military Hydrology, Report 10, Assessment and Field Examples of Continuous Wave Electromagnetic Surveying for Ground Water," Misc Paper EL-79-6, CEWES.
9. Butler, D.K., and Fitterman, D.V. 1986. "Transient Electromagnetic Methods for Ground Water Assessment," Misc Paper GL-86-27, CEWES.
10. Llopis, J.L., and Sjostrum, K.J. 1989. "Geophysical Investigation at Hazardous Waste Management Site 16, Radford Army Ammunition Plant, Radford, Virginia," Misc Paper GL-89-18, CEWES.

CALDWELL

The Impact of Atropine Sulfate on Flight
Performance, Vision, Tracking, and Electroencephalographic
Activity of Army Helicopter Pilots

J. A. Caldwell, Jr.
U. S. Army Aeromedical Research Laboratory
Fort Rucker, AL   36362

Introduction

Studies assessing the impact of doctrinal doses of atropine sulfate on the performance of Army helicopter pilots were undertaken by the U. S. Army Aeromedical Research Laboratory (USAARL) in 1984. The research has provided results of considerable relevance to operational aviation units by offering insight into the potential impact of a doctrinally approved antidote to chemical poisoning.

Atropine sulfate is an acetylcholine blocker which has been fielded in 2 mg autoinjectors for use as an antidote to nerve agent poisoning. Since chemical nerve agents consist primarily of acetylcholinesterase inhibitors which produce an excess of acetylcholine at receptor sites, drugs like atropine are a logical treatment choice.[1]

Mild exposure to anticholinesterase poisons produces symptoms such as anorexia, headache, weakness, and reduced visual acuity; whereas, moderate to high levels of exposure might be expected to produce symptoms ranging from bradycardia, nausea, and increased salivation, to respiratory difficulty, heart block, convulsions, and death. Immediate and adequate administration of atropine is a top priority after organophosphate poisoning has occurred.[2] U.S. Army personnel thus are trained to recognize the symptoms of nerve agent poisoning and administer appropriate atropine therapy (up to 6 mg) in a timely fashion.[3]

Although there is no question that atropine is an effective antidote to organophosphate poisoning, there have been concerns over the manner in which atropine may affect performance when it is injected because a soldier mistakenly assumes he or she has been poisoned. On the battle-field where there is a great deal of noise, smoke, and anxiety, it is conceivable that personnel could inaccurately perceive the presence of a chemical threat and initiate atropine therapy. The consequences of such action should be explored, particularly in an aviation context where the slightest performance impairment may produce dire consequences.

CALDWELL

Many of the effects of unchallenged atropine have been summarized by Lobb et al. (1985).[1] These include possible reduced alertness and increased anxiety, inhibition of sweat secretion, degraded visual acuity, photophobia, and potential impairments in central nervous system functioning. Research has documented atropine-related impairments in vision,[4] tracking,[5] and cortical activation.[6] In addition, investigators have identified detrimental effects of atropine on flight performance in fixed-wing[7] and rotary-wing simulators.[8]

The USAARL simulator study was the precursor to the in-flight atropine study presented here.[8] Study in the simulator determined few decrements attributable to the 2-mg dose of atropine, whereas the 4-mg dose degraded several aspects of helicopter pilot performance. Specifically, there were accuracy decrements during straight-and-level flight, a climbing turn, and (marginally) an instrument landing system approach which were primarily due to the 4-mg dose. The study showed that aviator performance would not "fall apart" because of atropine, but there would be decrements. To determine the full extent of these decrements, an in-flight evaluation was proposed.

The present investigation reports the effects of 2 mg and 4 mg of atropine sulfate on the actual in-flight performance of U.S. Army helicopter pilots. Vision, psychomotor tracking, and the electroencephalogram (EEG) are examined as well.

Method

Subjects

Twelve aviators in good health, possessing uncorrected 20/20 vision and normal hearing were used as subjects. Prior to enrollment, each participant was required to submit to a rigorous physical examination, including a cardiac stress test and an atropine sensitivity test.

Apparatus

Vision testing was conducted with an array of standard diagnostic tests for acuity, heterophoria, accommodation, stereopsis, and pupil diameter. Acuity was measured with a Snellen eye chart (either projected or hand-held miniature). Heterophoria, a test of whether the visual axes of the eyes are parallel, was measured with the Armed Forces vision testing apparatus. Accommodation was measured with a Prince rule and an accommodation target. Stereopsis was assessed with the TNO test. Pupil diameter was determined by measurements with a millimeter ruler.

Electroencephalographic (EEG) data were collected with a Cadwell Spectrum 32 brain mapper which provided for acquisition, storage, and analysis of 21 channels of EEG data. The device also generated the checkerboard patterns which were used to elicit visual evoked responses.

CALDWELL

In addition, a Cadwell 7400 evoked response system was used to generate stimuli and collect a single channel of data for an oddball (P300) paradigm.

Psychomotor tracking ability was assessed with a Zero Input Tracking Analyzer (ZITA) which presented a laterally moving cursor on a dot matrix display. A joystick was used to place the cursor over a centered target. The ZITA was interfaced to an Apple II+ microcomputer which controlled task difficulty and recorded data.

Flight performance evaluations were conducted in a UH-1H utility helicopter equipped with specialized components for recording instrument readings and pilot control movements while sequencing through a series of standardized flight maneuvers (Figure 1). Measures of altitude, airspeed, heading, pitch, roll, vertical speed, and instrument landing system (ILS) localizer and glide slope were collected.[9] Data were recorded on removable tape cartridges which were brought into the Laboratory at the conclusion of each flight. All of the relevant flight data along with safety-pilot indicators of start/stop points for each maneuver were transferred to a Digital Equipment Corporation (DEC) VAX 11/780 computer for subsequent analyses.
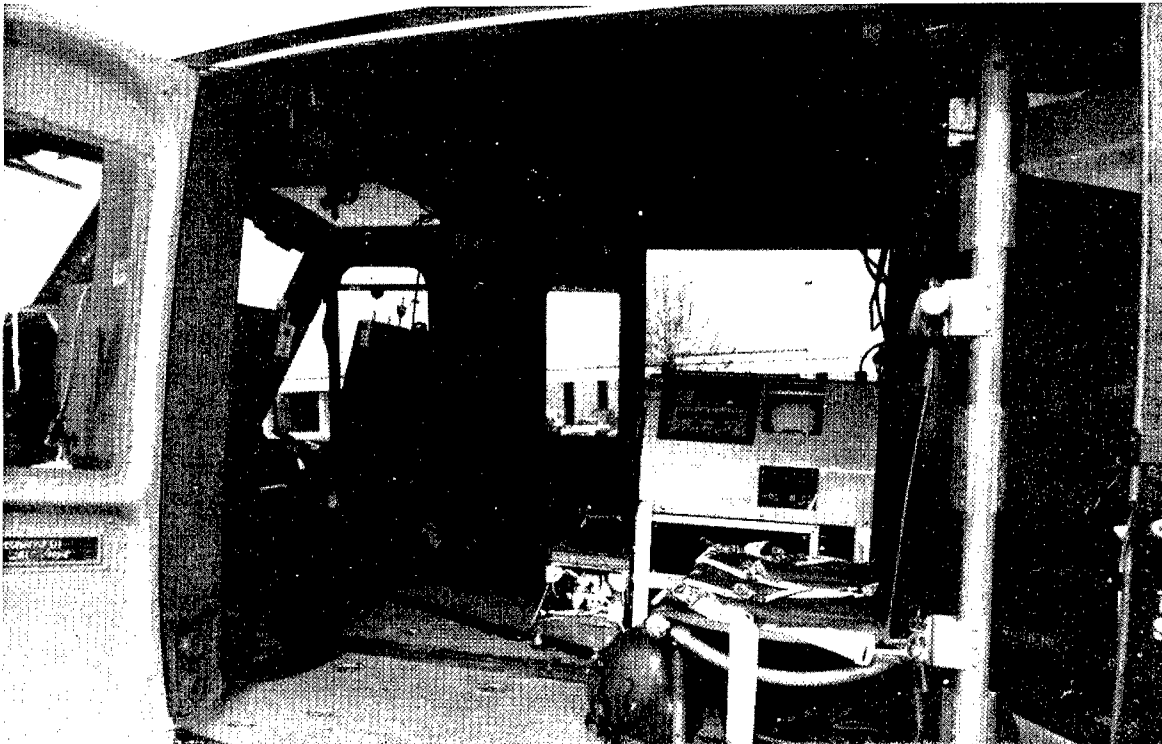

Figure 1. Interior of the UH-1 equipped with the computerized in-flight monitoring system.

Procedure

Subjects remained in the Laboratory for periods lasting up to 11 days during which they participated in a variety of testing sessions. The first 2-3 days were used to train participants on the proper execution of the flight profile along with concurrent training on laboratory tests. On the evening of a subject's first training day, 25 EEG electrodes were attached for EEG recording.

Training on the flight profile and laboratory tests was not considered complete until computerized assessments of flight performance indicated the subject had reached asymptotic levels. Once this occurred, the next day was designated as the beginning of the counterbalanced, double-blind dose sequence.

There were three dose days, each separated by a control day to allow adequate atropine clearance. On each dose-administration day, only one dose (placebo, 2 mg, or 4 mg I.M.) was given. Each dose day consisted of a morning session of laboratory tests, a single injection, a morning flight, a noon laboratory test session (beginning about 2.5 hours postdose), an afternoon flight, and an evening laboratory session (beginning about 8 hours postdose). Control days were similar except that there was no injection, no flying, and no evening test session.

The laboratory test sessions began with the vision tests of accommodation, stereopsis, heterophorias, acuity, and pupil diameter. These were followed by electrophysiological assessments which included visual evoked responses to a series of checkerboard patterns (ranging from 4 x 4 checks to 64 x 64 checks), a visual P300, and a 2-minute resting eyes open/eyes closed EEG. This was followed closely by three increasingly difficult tracking tests where subjects were required to hold a cursor over a centered tracking target using a joystick.

The two flights (on each dose day) were each approximately 2 hours long. The maneuvers flown were identical for every dose condition across all subjects. Performance was evaluated via the computerized aircraft monitoring system during higher altitude work consisting of five straight-and-level segments, two standard-rate level turns, a climbing turn, a descending turn, a straight climb and descent, and two steep turns, as well as during an "instrument only" straight-and-level segment and an instrument landing system (ILS) approach. There was also a portion of the profile which included confined-area operations and navigation (Figure 2), but these were not scored by the computer.

Thus, each dose day consisted of three laboratory testing sessions (the first of which was predose) interspersed with two flights in the UH-1H. Control days consisted of only two laboratory sessions and no flights.
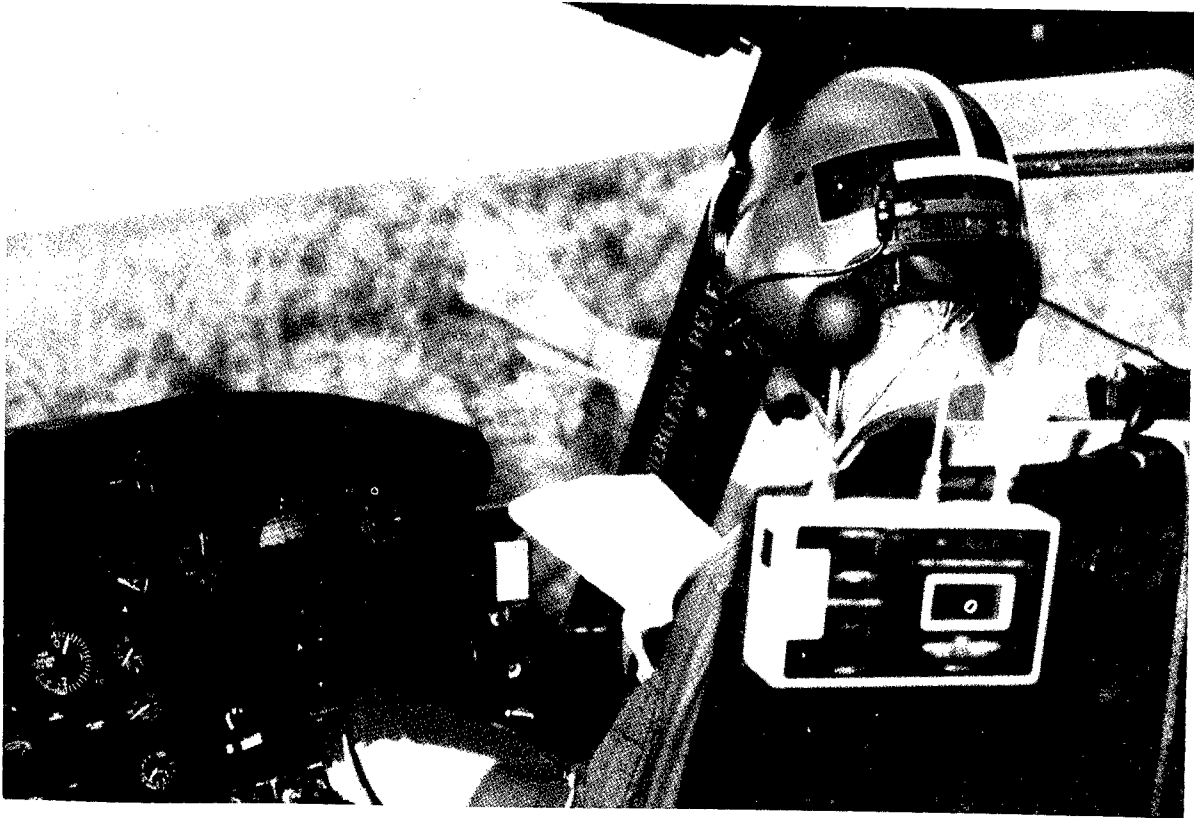
CALDWELL

Figure 2. A participant performs navigation after completing the
precision upper-altitude work and confined-area operations.

Results

Statistical procedures

All data were analyzed with BMDP4V, multivariate and univariate
analysis of variance/analysis of covariance.[10] The vision data and the
flight data were analyzed with analysis of variance, but the tracking
scores and EEG data were analyzed with analysis of covariance in which
data from the predose session was used as the covariate. The few
instances of missing data were handled with BMDPAM by substituting the
appropriate cell means. All data originally expressed as percentages
were transformed with the 2*arcsin(sqrt(x)) procedure recommended by
Winer (1971) prior to analysis.[11] Sphericity violations were handled
with Box/Geisser-Greenhouse corrected degrees of freedom.[12]

Because of the voluminous number of significant findings, only those
central to this presentation are reported here. A technical report which
describes the entire study is forthcoming.

Vision

Pupil diameter. Analysis of pupil diameter (in millimeters) revealed an interaction between dose and session for both eyes (right- $F_{(4,44)}= 17.42, p<.0001$; and left- $F_{(4,44)}=17.85, p<.0001$). This was because of increases in pupil size under both 2 mg and 4 mg of atropine at noon and in the evening ($p<.05$). Also, there were dose main effects which restated the above finding (Figure 3).

Accommodation. Accommodative power was expressed as the distance from the eye (in centimeters) at which a subject could focus an accommodation target. Examination of these data showed an interaction between dose and session (right eye- $F_{(1.36,15.00)}=11.47, p=.0022$; and left eye- $F_{(1.55,17.03)}=7.98, p=.0058$) which was due to reductions in accommodation under both 2 mg and 4 mg atropine at both postdose sessions (Figure 4). There was a similar dose main effect.

Stereopsis. The stereopsis data were expressed as the lowest angular disparity at which a subject was able to accurately determine the orientation of stereogram targets perceived in depth. Stereopsis is an indicator of depth perception. Analysis of log-transformed scores revealed a dose by session interaction ($F_{(2.07,20.75)}=5.50, p=.0114$) attributable to reductions in depth perception under both 2 mg and 4 mg of atropine at noon and in the evening ($p<.05$). Both dose conditions were different from placebo, but not from each other. There was also a dose main effect for the same reason (Figure 5).
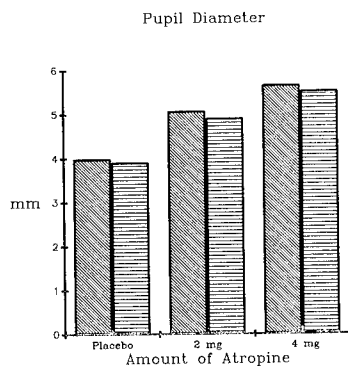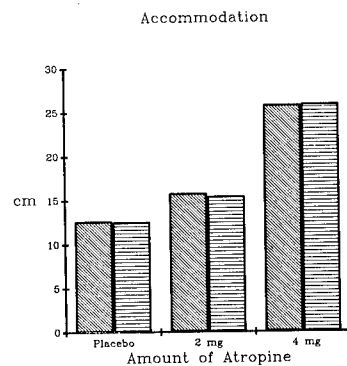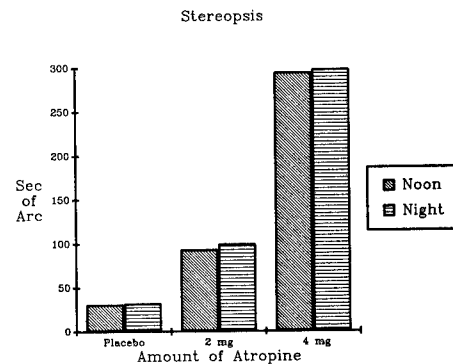

Figure 3.   Figure 4.   Figure 5.

Phorias. The scores for lateral phorias (near and far) depicted the extent to which the two eyes functioned in parallel along a lateral axis. The data for vertical phorias revealed the extent to which the two eyes travelled in parallel along a vertical axis. There were no significant changes in far phorias because of atropine. However, for near lateral phoria there was a dose by session interaction ($F_{(4,44)}=10.81, p<.0001$)

attributable to increasing esophoria (or tendency to become cross-eyed) under 2 mg and 4 mg atropine both at noon and in the evening. There was a dose main effect for the same reason (Figure 6). The data for near vertical phoria also produced a dose by session interaction $(F(2.29,25.15)=3.69,p=.0342)$. Further analysis revealed an increase in left hyperphoria (left eye looking slightly higher than the right) under 2 mg and 4 mg atropine only during the noon session (Figure 7). The dose main effect found here was due to the same type of effect except that collapsing the sessions resulted in a finding of increased left hyperphoria only under 4 mg.
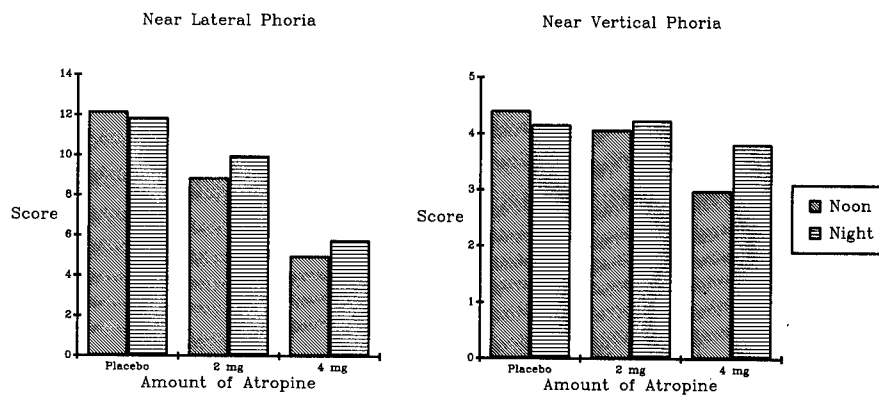
Near Lateral Phoria

Near Vertical Phoria

Figure 6.

Figure 7.

Acuity. The visual acuity data, while unsuited for parametric analysis, deserves mention. Results indicated that neither dose of atropine affected far visual acuity, but near visual acuity was impaired in 5 of the 12 subjects under 4 mg.

Electroencephalographic data

Resting EEG. Data collected from the four midline electrodes were subjected to power spectral analysis to determine the extent of cortical activation at each site. In general, a fast, desynchronous EEG trace is associated with arousal or activation, whereas a slow synchronous trace is associated with relaxation and/or deep sleep. The EEG activity is divided among four bands designated as delta (1.5-3.5 Hz), theta (3.5-7.5 Hz), alpha (7.5-12.5 Hz), and beta (12.5-20 Hz). The amount of activity falling within each band is expressed as a transformed percentage of total power at each electrode. The electrodes were placed in accordance with International 10-20 coordinates: Fz-frontal, Cz-central, Pz-parietal, and Oz-occipital.

There were significant atropine-related increases in **delta** activity at Fz $(F(2,21)=6.35,p=.007)$, Cz $(F(2,21)=4.01,p=.0335)$, and Pz

(F(2,21)=7.25,p=.004). In all three cases, there was more slow-wave activity under 4 mg than under placebo (p<.05), and at Fz there was more slow-wave under 2 mg as well (Figure 8).

Theta activity was unaffected by atropine at both Fz and Cz, but there were increases in theta under 4 mg compared to 2 mg and placebo at Pz (F(2,21)=5.50,p=.0120) and Oz (F(2,21)=7.22,p=.0041). These effects are depicted in Figure 9. Also, there was a dose by session interaction (F(2,22)=13.41,p=.0002) in the amount of theta at Oz which was because of a stronger atropine effect at noon than in the evening (p<.05).

Delta Activity   Theta Activity
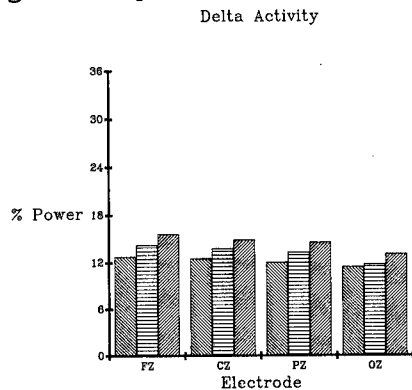

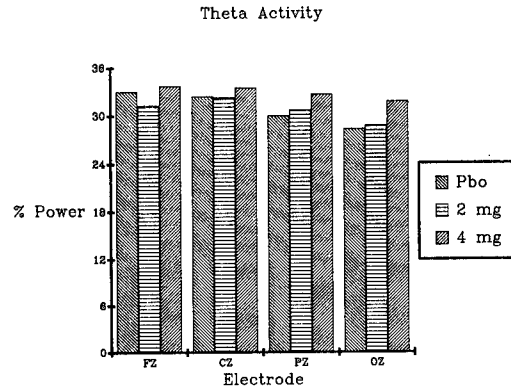Figure 8.   Figure 9.

The relative power of alpha activity was substantially reduced by 4 mg of atropine at Fz (F(2,21)=5.32,p=.0136), Cz (F(2,21)=3.59, p=.0457), Pz (F(2,21)=7.95,p=0.0027), and Oz (F(2,21)=7.76,p=.0030). There were not significant reductions from placebo to 2 mg at any of these electrodes (Figure 10).

There were no dose main effects with regard to the relative power of beta activity at any electrode (Figure 11). However, there was a dose by session interaction at Oz (F(2,22)=3.82,p=.0376) which tended (p=.057) to be because of substantially lower beta under 4 mg relative to placebo during only the noon session.
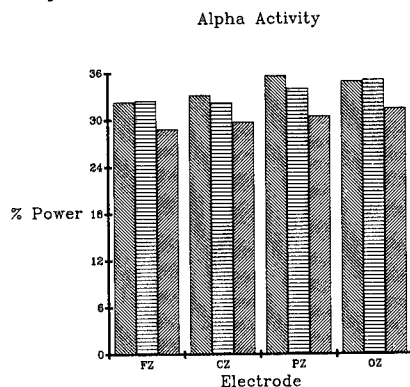
Alpha Activity   Beta Activity

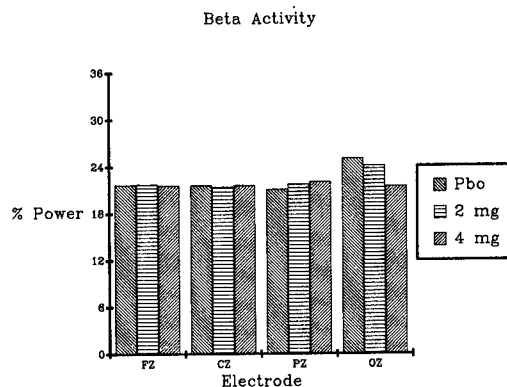Figure 10.   Figure 11.

Evoked responses. In addition to the resting EEG, data were collected in two types of evoked response tasks. The first type of task, in which visual presentations of reversing checkerboard patterns were viewed while EEG was being sampled, was administered to gain some insight into how atropine affected stimulus registration. The second type of task, in which a checkerboard pattern reversed only 40 out of 200 sweeps and data were collected in response to both nonreversals and reversals, was used to gain some insight into central nervous system processing speed. Both these tasks yielded cortical evoked responses scored from a single lead (Oz in the first and Pz in the second), in terms of latencies and amplitudes of various peaks (thought to correspond to aspects of stimulus recognition and cognitive processing). One subject's data from the first task was eliminated because he was an outlier.

Analysis of data from the first task (simple checkerboard reversals) showed an interaction between dose and session $(F(2,20)=3.48,p=.0506)$ for the negative peak at 75 ms poststimulus (N75). This effect was because of reductions in amplitude under 4 mg as compared to 2 mg and placebo $(p<.05)$ at the noon session only (Figure 12). The dose main effect was basically the same. Also, there was a dose by session interaction $(F(2,20)=7.11,p=.0046)$ for the positive peak at 100 ms poststimulus (P100) which was due to increases in amplitude under 4 mg and 2 mg $(p<.05)$ in the evening only (Figure 13). The dose main effect was quite similar. The latencies of these components were not altered by atropine.
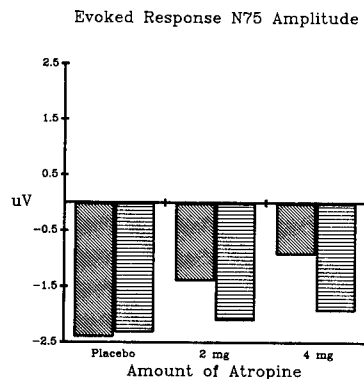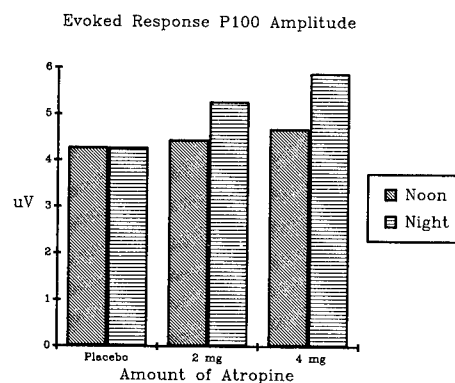

Figure 12.

Figure 13.

Analysis of data from the second task (requiring subjects to attend to infrequent pattern reversals) revealed a dose-related reduction $(F(2,15)=3.76,p=.0474)$ in the positive peak occurring at 300 ms post-stimulus (P300). Amplitude was reduced under the 2- and 4-mg conditions

CALDWELL

relative to placebo (p<.05). Also, there was a dose-related increase of the P300 latency (F(2,15)=5.79,p=.0137) under 4 mg as compared to 2 mg or placebo (p<.05). See Figures 14 and 15.
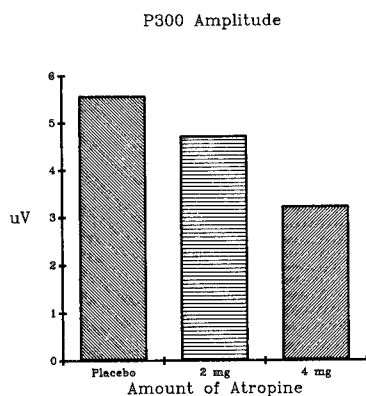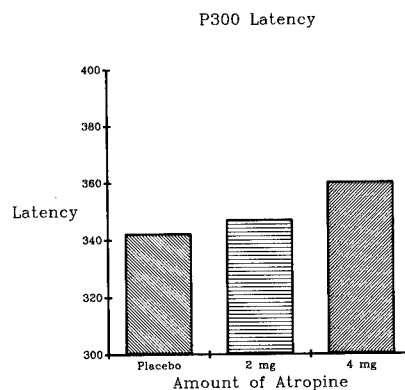
P300 Amplitude

P300 Latency

Figure 14.

Figure 15.

Psychomotor tracking

There were three levels of tracking difficulty. Each task also contained an auxiliary auditory distraction component, but this will not be discussed here. Subjects began with the easiest task in which the cursor (to be held over the target) responded rapidly to any joystick movement. The remaining two tasks became increasingly difficult because of increases in the amount of effective delay from joystick input to cursor response. The scores from each task were computer-generated deviations of the cursor from the target (a time and distance integration). Perfect tracking would have resulted in a score of 0 while the worst possible outcome was a score of 9999.

Tracking level 1. Analysis of tracking deviation scores on the easiest tracking task revealed a dose main effect (F(1.05,10.98)= 7.03,p=.0216) attributable to poorer performance under 4 mg than under the other two conditions (p<.05). See Figure 16.

Tracking level 2. On this intermediate task, the dose by session interaction was significant (F(1.01,11.06)=4.89,p=.0489) because of atropine-related tracking changes during only the noon session. Here, there was a tendency (p=.06) for tracking to decline under 4 mg as compared to placebo (Figure 17).

Tracking level 3. Atropine effects on the most difficult task likewise were detected only during the noon session. Examination of the dose by session interaction (F(2,22)=3.52,p=.0473) revealed poorer performance under 4 mg than under 2 mg or placebo during only the noon session (p<.05). This effect is depicted in Figure 18.

ZITA
Task 1

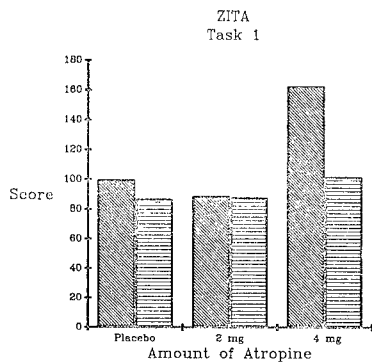Figure 16.

ZITA
Task 2

Figure 17.

ZITA
Task 3

□ Noon
□ Night
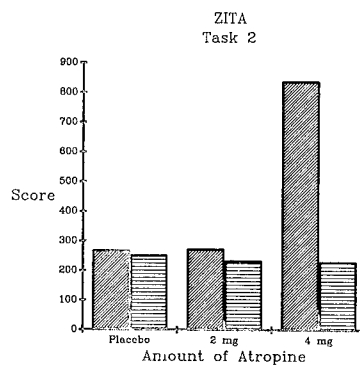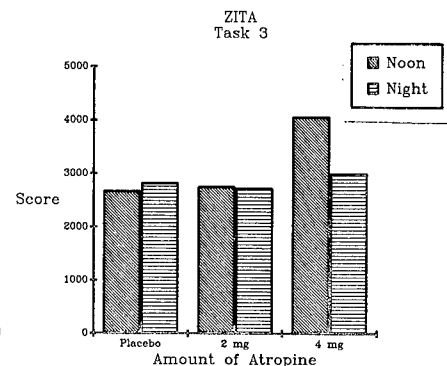
Figure 18.

Flight performance

For each maneuver, the flight performance of each subject was expressed in terms of scores which ranged from 0 to 100. Subjects were graded on the control of specific parameters such as heading, altitude, and airspeed which were supposed to be maintained according to instructions. For instance, during a straight-and-level maneuver, a subject might have been told to maintain a heading of 180 degrees, an altitude of 1000 feet, and an airspeed of 90 knots. If the subject was able to perfectly adhere to these standards for the entire maneuver, he earned a score of 100 on each parameter. Scores of less than 100 resulted from control deviations of various magnitudes. Flight scores from each of the flights under placebo, 2 mg, and 4 mg were transformed and then analyzed in a series of repeated measures analyses of variance. The factors were dose (placebo, 2 mg, and 4 mg), flight (morning and afternoon), and maneuver (as in 5 straight-and-levels). Several dose-related performance effects are depicted in Figure 19.
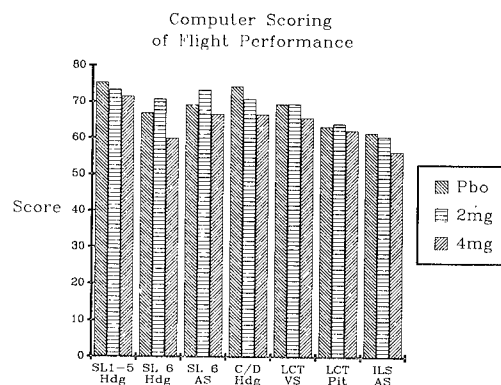

Computer Scoring
of Flight Performance

□ Pbo
□ 2mg
□ 4mg

Figure 19.

Straight and level. There were several straight-and-level segments used in the profile. The first five (SLs 1-5) were conducted with visual references, and the last one (SL 6) was conducted using instrument references only. Across SLs 1-5 there was an atropine-related reduction in the accuracy of heading control ($F(2,22)=3.67, p=.0421$) under the 4-mg dose compared to placebo. This finding is important since it indicates consistent atropine effects across all five SLs and it suggests there were atropine-related decrements as early as about 17 minutes postdose. A similar effect was also seen much later in the flights during SL 6 ($F(2,22)=7.22, p=.0039$) where the 4-mg dose produced lower heading scores than 2 mg or placebo ($p<.05$). In addition, there was a dose effect on airspeed control during SL 6 ($F(2,22)=8.54, p=.0018$) which resulted from poorer performance under placebo and 4 mg compared to 2 mg ($p<.05$). Also, there was a dose by flight interaction on altitude control ($F(2,22)=3.90, p=.0354$) which was apparently due to a performance improvement from morning to afternoon only under placebo ($p<.05$).

Standard-rate turns. There were no dose main effects on the right and left level turns; however, there was an interaction among dose, flight, and maneuver for vertical speed control ($F(2,22)=3.90, p=.0355$). Subsequent analysis indicated better morning than afternoon performance on the first turn under the 4 mg dose -- an effect probably due to the amount of time it took for atropine to exert profound effects on this type of maneuver. The first turn was only 14 minutes into the flight.

Straight climb and descent. Performance on these two maneuvers was affected with regard to heading control ($F(2,22)=6.00, p=.0083$). Control was poorer under 4 mg than under placebo ($p<.05$).

Steep turns. Analysis of performance during the two steep turns revealed one atropine-related effect which was an interaction between dose and flight ($F(2,22)=3.57, p=.0456$). This was due to better morning than afternoon control of bank angle under the 4-mg dose ($p<.05$). This may have been due to insufficient time from dose to these maneuvers (31-38 minutes in the morning) for atropine to have made a substantial impact on such difficult maneuvers where there was probably already a great deal of non-drug-related variance.

Descending and climbing turns. Examination of these two maneuvers revealed dose by maneuver interactions on vertical speed control ($F(2,22)=4.85, p=.0180$) and pitch control ($F(2,22)=3.94, p=.0344$). Both of these were found to be attributable to atropine effects during only the climbing turn (approximately 52 minutes into each flight) where 4 mg produced poorer performance than the other two conditions ($p<.05$ for vertical speed comparisons and only marginal significance for pitch comparisons).

Instrument landing system. Performance during the instrument approach to landing also was found to be affected by atropine in that airspeed control ($F(2,22)=6.10, p=.0078$) suffered under the 4-mg condition ($p<.05$).

Discussion

The first investigation of the effects of 2 mg and 4 mg atropine sulfate in an actual in-flight helicopter environment revealed a number of interesting effects. These were bolstered by laboratory tests which contributed additional operationally relevant information.

The findings that atropine increased pupil diameter, decreased accommodative power, and reduced visual acuity agreed with the work of others, summarized by Headley (1982).[4] Added to these effects, we found evidence of impaired depth perception and double vision which would further interfere with performance on visual tasks. With the exception of near visual acuity (reduced only under 4 mg), statistically significant vision impairments were observed after both the 2- and 4-mg doses of atropine.

Changes in the resting electroencephalogram presented evidence for reduced cortical activation which was congruous with the research summarized by Longo (1966).[6] That author reported atropine-induced, consistent slowing of the EEG accompanied by substantial alpha reductions under atropine. These effects were clearly obtained with the 4-mg dose in this study and this is important since it indicates drowsiness and decreased alertness which may cause problems for soldiers, especially pilots, during tactical scenarios. The findings with regard to cortical-evoked responses tended to support the conclusion that atropine was causing some sedation (in addition to general vision problems). The reduced amplitude and increased latency of the P300 component suggest atropine-related impairments with attention and/or the speed of mental processing.

The reductions in tracking accuracy found across three levels of tracking difficulty were compatible with earlier findings. Penetar and Beatrice (1986),[5] testing at 150 minutes postdose, discovered the same sort of dose relation: the 4-mg dose consistently produced decrements while 2 mg failed to make a significant impact. Whereas Penetar and Beatrice observed this effect at 2.5 hours postdose (their last test session),[5] we saw the same problem as late as about 4 hours postdose. However, by the evening session (approximately 9.5 hours postdose), this effect of the larger dose was gone.

Most important, impairments in the accuracy of flight performance under atropine also were observed. Especially when considering the impact of the 4-mg dose, our findings regarding accuracy problems on maneuvers such as straight-and-levels, turns, climbs and descents, and instrument landing system approaches lend support to the results of Dellinger et al. (1987)[7] and Simmons et al. (1989).[8] Both of these investigations reported 4-mg atropine effects which occurred early (within 1 hour postdose) and were fairly persistent (remaining at 3 hours or beyond) across several maneuvers. Our study suggests some flight performance decrements under 4 mg remained as long as 7 hours or more

postdose. The 2-mg dose was typically far less powerful. Our results provide information beyond what was obtained in the earlier studies by documenting the characteristics of atropine effects in an actual in-flight (as opposed to simulator) environment.

In summary, it appears a large dose (4 mg) of atropine sulfate will result in a number of vision impairments and performance decrements which may begin within 30 minutes postdose and persist for as long as 7 to 8 hours. Also, there likely will be atropine-related drowsiness lasting 8 hours or more postdose. Based on our tracking task, performance begins to return to normal (after 4 mg atropine) by about 9 hours postdose. Visual inspection of some of the data frequently suggested slight decrements under the 2-mg dose, but these were typically not statistically significant with regard to any indicators other than tests of vision.

The potential impact of atropine in an operational aviation scenario is difficult to estimate. Some of the vision data suggest pilots may have problems with map and instrument readings, whereas the tracking data imply less responsiveness to task demands under 4 mg atropine. The EEG data show atropine exerts sedative effects which may lead to reduced alertness and vigilance, particularly under the larger dose. All of this information, combined with the analysis of flight performance, suggests a helicopter pilot mistakenly could inject up to 4 mg atropine and still safely return his helicopter to base <u>if</u> he is not required to handle serious in-flight emergencies, perform overly-taxing secondary tasks, or execute maneuvers which require very precise aircraft control (i.e., formation flight or confined-area operations). However, his performance should be considered seriously impaired for several hours after atropine administration. Also, while the data collected under the smaller dose presents only limited cause for concern, even 2 mg of atropine may produce effects similar to low levels of alcohol for which "12-hours-from-bottle-to-throttle" is required even after consumption of a single alcoholic beverage.[13]

References

1. Lobb, M. L., Phillips, J. D., and Winter, A. S. (1985). <u>Effects of atropine sulfate on aircrew performance: A review and evaluation</u>. Brooks Air Force Base, TX: U.S. Air Force School of Aerospace Medicine: Technical report no. 85-48.

2. Mackey, C. L. (1982). Anticholinesterase insecticide poisoning. <u>Heart and lung</u>, <u>11(5)</u>, 479-484.

3. Department of the Army (1988). <u>First aid for soldiers</u>. Washington, DC: U.S. Army Field Manual FM 21-11.

4. Headley, D. B. (1982). Effects of atropine sulfate and pralidoxime chloride on visual, physiological, performance, subjective, and cognitive variables in man: A review. Military medicine, 147, 122-132.

5. Penetar, D. M., and Beatrice, E. S. (1986). Effects of atropine on human pursuit tracking performance. Aviation, space, and environmental medicine, 57, 654-658.

6. Longo, V. G. (1966). Behavioral and electroencephalographic effects of atropine and related compounds. Pharmacological Review. 18(2), 965-996.

7. Dellinger, J. A., Taylor, H. L., and Porges, S. W. (1987). Atropine sulfate effects on aviator performance and on respiratory-heart period interactions. Aviation, space, and environmental medicine, 58, 333-338.

8. Simmons, R. R.. Caldwell, J. A., Stone, L. W., Carter, D. J., Behar, I., Mitchell, G. W., Knox, F. S., Jones, H. D., and Taylor, P. L. (1989). Effects of the chemical defense antidote atropine sulfate on helicopter pilot performance: A simulator study. Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory: Technical report no. 89-17.

9. Mitchell, A., Lewis, A., Jones, H., Higdon, A., and Baer, D. (1988). Aircraft in-flight monitoring system (AIMS). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory. USAARL Letter report 88-12-5-2.

10. Dixon, W. J., Brown, M. B., Engleman, L., Frane, J. W., Hill, M. A., Jennrich, R. I., and Toporek, J. D., eds. (1983). BMDP statistical software. Berkeley: University of California Press.

11. Winer, B. J. (1971). Statistical principles in experimental design. 2nd. ed. New York: McGraw-Hill.

12. Grieve, A. P. (1984). Tests of sphericity of normal distributions and the analysis of repeated measures designs. Psychometrika, 49(2), 257-267.

13. Department of the U.S. Army. (1986). Temporary flying restrictions due to exogenous factors. Washington, DC: U.S. Army Regulation No. 40-8, para 4a(1).

The Morphology of Polymer Chains
by Transmission Electron Microscopy (U)

*Dr. Yvon P. Carignan
U.S. Army Armament Research, Development and Engineering Center
Picatinny Arsenal, NJ 07806-5000

In spite of all the recent successes reported in the study of polymer shapes and dimensions based on the techniques of small angle neutron scattering (SANS), small angle x-ray scattering (SAXS), and small angle light scattering (SALS), there are serious limitations in their scope (1). For instance only average values, such as root-mean-square end-to-end distance and radius of gyration of a polymer chain can be extracted from these techniques. In other words, these techniques cannot provide information on the actual conformation or dimension of the individual polymer chain. Also since all three techniques are indirect, the analysis of the raw data is not straightforward and requires various manipulations and faith in existing theories to draw conclusions. Finally with regard to SANS, it must be noted that only a limited number of facilities in the U.S. have this capability, thus restricting its use.

In the present work we have taken the challenge of developing a direct technique which could provide the ultimate information on the actual morphology and dimensions of the individual polymer chains. We were convinced that if successful, an experimental technique which would give such detailed information on polymer chains would not only be received by researchers and technologists as a major breakthrough towards the simplification of their efforts in understanding and improving the performance of current and future polymeric materials but would also sort out the best, most realistic theory and model among the many being offered today to explain the properties of polymers.

Towards this objective, we have turned to the concept of mass thickness contrast microscopy, which could be exploited by tagging the polymer under study with a heavy atom, and after forming thin films of the tagged polymer dispersed in a matrix of untagged polymer, perform examination of the films by Transmission Electron Microscopy (TEM). Polystyrene (PS) was chosen as the polymer mainly because of its

availability over a wide range of molecular weights (M_W= 1x10^4 to 2x 10^6) and very low polydispersity (M_W/M_N<1.1). Furthermore, PS has been over the years more scrutinized by both theoreticians and experimentalists than any other polymer in terms of its properties, conformation and dimensions. To enhance the contrast against the electron beam, the PS molecular chains were tagged with iodine atoms along the chain and these tagged PS chains were mixed with the starting PS for observation by TEM in the form of ultrathin films. Obviously, the visualization of the individual tagged polymer chains would be most appealing and convincing. An encouraging note as to the prospective success of our approach came from the work of Furuya et al (2). These authors tagged a number of short rigid rodlike molecules at both ends with metal clusters to enhance contrast. Using scanning electron microscopy they could observe paired spots from the contrast produced by the electron-dense metal clusters which were separated by distances which closely approximate the lengths of the rodlike molecules. We were nevertheless aware of the dangers for misinterpretation inherent to this type of approach. As pointed out by E. L. Thomas in his critical review on the "Transmission Electron Microscopy of Polymers" (3), a number of papers have appeared in the literature which have completely misinterpreted the images seen via mass thickness contrast after tagging with a heavy atom.

EXPERIMENTAL

SYNTHESIS OF POLY (P-IODOSTYRENE). Poly (p-iodostyrene) or iodinated polystyrene (IPS) samples were prepared by iodination of polystyrene using the procedure described by Braun (4). Polystyrene dissolved in nitrobenzene was reacted with the mixture of iodine and iodic acid. The starting polystyrenes were polystyrene standards with narrow molecular weight distribution (M_W= 2x10^6; M_W/M_N = 1.10). After the reaction was terminated, the polymer was recovered by precipitation into methanol. The products were then dissolved in a solvent mixture (benzene/methanol 90/10 v/v), freeze-dried, and dried under vacuum at room temperature for a week to remove the residual solvent. The iodine content of the samples was determined by elementary analysis. The relationship between the iodine content of the product and reaction time is shown in Fig. 1. The iodine content increases monotonically with increasing reaction time, as was pointed out by Braun. Also as shown in the figure, the molecular weight of the starting polystyrene (PS) has no effect on the rate of iodination reaction. Therefore, we can use this calibration curve for preparing materials with the desired iodine content for PS of any molecular weight. Through viscosity (4) and light scattering measurements (8) it was established that the

iodination reaction does not change the molecular weight of the starting polymer.

TRANSMISSION ELECTRON MICROSCOPY (TEM). The TEM study of the polymer films was conducted with a JEOL 100 CX-II microscope operating at 100 kV. The method of observation was based on the scattering contrast mechanism (5,6). As represented schematically in Fig. 2, the atomic scattering factor or the number of electrons scattered at a specific angle α is directly related to the atomic number (Z) of the scattering atom. Therefore, polymer chains containing a heavy atom, such as iodine with Z=53 will scatter more electrons than the surrounding chains containing only light atoms (C,H,N...). A contrast will be obtained by blocking all the electrons scattered past a critical angle which is determined by the radius of the objective aperture (R) and the focal length of the objective lens (f) as shown in Fig. 3. Optimum contrast will be obtained for large f and small R. In this investigation the following parameters were used: f = 4.5mm and R = 10 μm with iodine as the heavy atom.

THIN FILM SPECIMEN PREPARATION. Iodinated polystyrene (IPS) and polystyrene (PS) were dissolved separately in toluene at 0.5w% concentration. These two solutions were mixed to obtain a ratio of IPS to PS ranging from 1:1 to 1:100. For ratios below 1:100 the solution of IPS was further diluted by a factor of ten before mixing with the PS solution. The final solutions were stirred overnight before preparing the films.

Thin film samples were cast on a glass slide and dried in a bell jar containing toluene vapor to decrease the evaporation rate. Thin films (about 1000Å) were formed in two hours. After solvent removal the films were cut on the glass slide into 2mm x 2mm squares before being dipped into distilled water. The thin films were picked up by copper grids and dried under vacuum for at least twenty-four hours before examination by TEM.

RESULTS AND DISCUSSIONS

The TEM results for an IPS:PS concentration ratio of 1:10 are given in Fig. 4. One observes a number of dark gray spots on an otherwise structureless gray background. We interpret these gray spots as representing the projection into 2-D space of the spherical volume occupied by random coiled IPS chains. This being the case, it would follow that reducing the (IPS:PS) ratio should correspondingly reduce the number (density) of spots on the photomicrographs. Indeed, as listed in Table I the number of spots decreases accordingly, thus

strengthening the argument that indeed we are observing 2-D projections of individual polymer chains of IPS.

TABLE I.   Density of Dark Grey Spots as a Function of IPS:PS Ratios

IPS:PS	Spots Count
1:10	40
1:20	20
1:50	7

By actually measuring the diameter of each spot on the photomicrograph (N=60) we have constructed the histogram shown in Fig. 5. A rough estimate of the average diameter of the spots is $650\text{\AA}$. Comparing this value with the value of $950\text{\AA}$ calculated for PS of this molecular weight ($M = 2 \times 10^{7}$) in the unperturbed state, based on the random coil model suggests either; a) that our measurements of the gray spots are underestimating the actual size of the spots simply because the contrast within the spots gradually decreases from the center to the periphery thus making estimation of the true boundary of the spots very difficult; b) the IPS chains achieve a more compact coil configuration relative to the PS chains. However, previous work with IPS in the bulk using x-ray scattering (7,8) contradicts this hypothesis. This work has shown that the radius of gyration of IPS is in fact 6% greater than that of PS. Therefore, an increase in dimensions of IPS relative to PS is predicted in contrast to the 30% decrease observed in the present work; c) our results may simply confirm the theoretical predictions of Lindenmeyer (9), i.e., molecules exist in the more densely coiled conformations than the random coiled form.

We are currently investigating the merits of suggestion (a), i.e. our measurements are underestimating the actual size of the polymer simply because of the fading contrast and loss of sharpness at the boundary of the TEM image.

A number of other IPS:PS ratios were studied with IPS containing both 40% and 20% iodine with the results shown in Table II. For the materials with high iodine content (40%) we note that at high ratio of IPS to PS (1:1) there is evidence of aggregation/phase separation of the IPS chains. At ratios of 1:10 to 1:50 one can observe the individual polymer chains which appear as grey circular spots on a lighter background. As mentioned before these circular spots are interpreted as being the two dimensional projection of spherical random coil chains. At ratios of the order of 1:100 or smaller the number density or concentration of IPS molecules is too small to show the

presence of a single IPS molecule within the dimensions covered in the photomicrographs. For materials containing half the number of iodine atoms (20%) similar results are obtained except for the weaker more diffuse contrast achieved.

TABLE II - Summary of TEM Results

Iodine Content	Mixing Ratios IPS:PS	Observations
40%	1:1	Phase Separated
	1:10	Random Coil
	1:20	Random Coil
	1:50	Random Coil
	1:100	No Structure
	1:1000	No Structure
20%	1:1	Phase Separated
	1:20	Random Coil
	1:50	No Structure
	1:100	No Structure
	1:1000	No Structure

CONCLUSIONS

This study provides for the first time direct evidence on the dimensions, conformation and shapes of individual polymer chains and the occurrence of chain aggregation under a given set of conditions. However this effort is essentially preliminary and further experiments are being conducted to test the scope of the technique. It is evident that the evolution of the progress to be achieved in the development of this unique technique must be of great interest in regards to the Army Research and Development mission considering the major and increasing reliance being placed on the variety of polymers and copolymer compositions used in a large variety of end items.

ACKNOWLEDGMENTS

Masanori Hara and the graduate students M. Bellinger, P. Jar and Y. Wang.

LITERATURE CITED

1. Higgins, J. S.; and Stein, R. S. J. Appl. Cryst., 1978, 11, 346.

2. Furuya, F. R.; Miller, L. L.; Hainfeld, J. F.; Christopfel, W. C.; Kenney, P. W. J. Am. Chem. Soc., 1988, 110, 641.

3. Thomas, E. L.; Structure of Crystalline Polymers; Hall, I. ed.; Elsevier Applied Science: New York, 1984; Chapter 3.

4. Braun, D.; Makromol. Chem., 1959, 30, 85.

5. Reimer, L.; Transmission Electron Microscopy; Elsevier Applied Science: New York, 1988: Chapter 3.

6. Henkee, C. S.; Transmission Electron Microscopy; Elsevier Applied Science: New York, 1988; Chapter 3.

7. Hayashi, H.; Hamada, F.; Nakajima, A.; Macromolecules, 1974, 7, 959.

8. Hayashi, H.; Hamada, F.; Nakajima, A.; Macromolecules, 1976, 9, 543.

9. Lindenmeyer, P. H.; J. Macromol. Sci., Phys., 1973, B8(1-2), 361.

REACTION CURVE
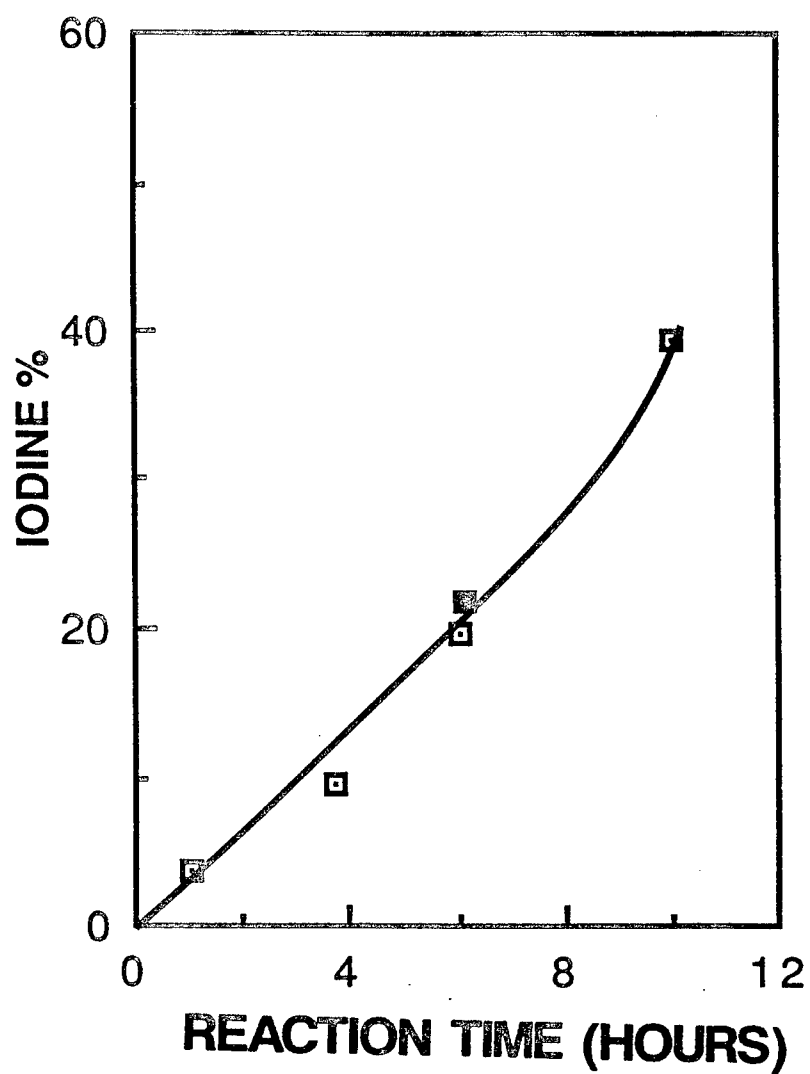
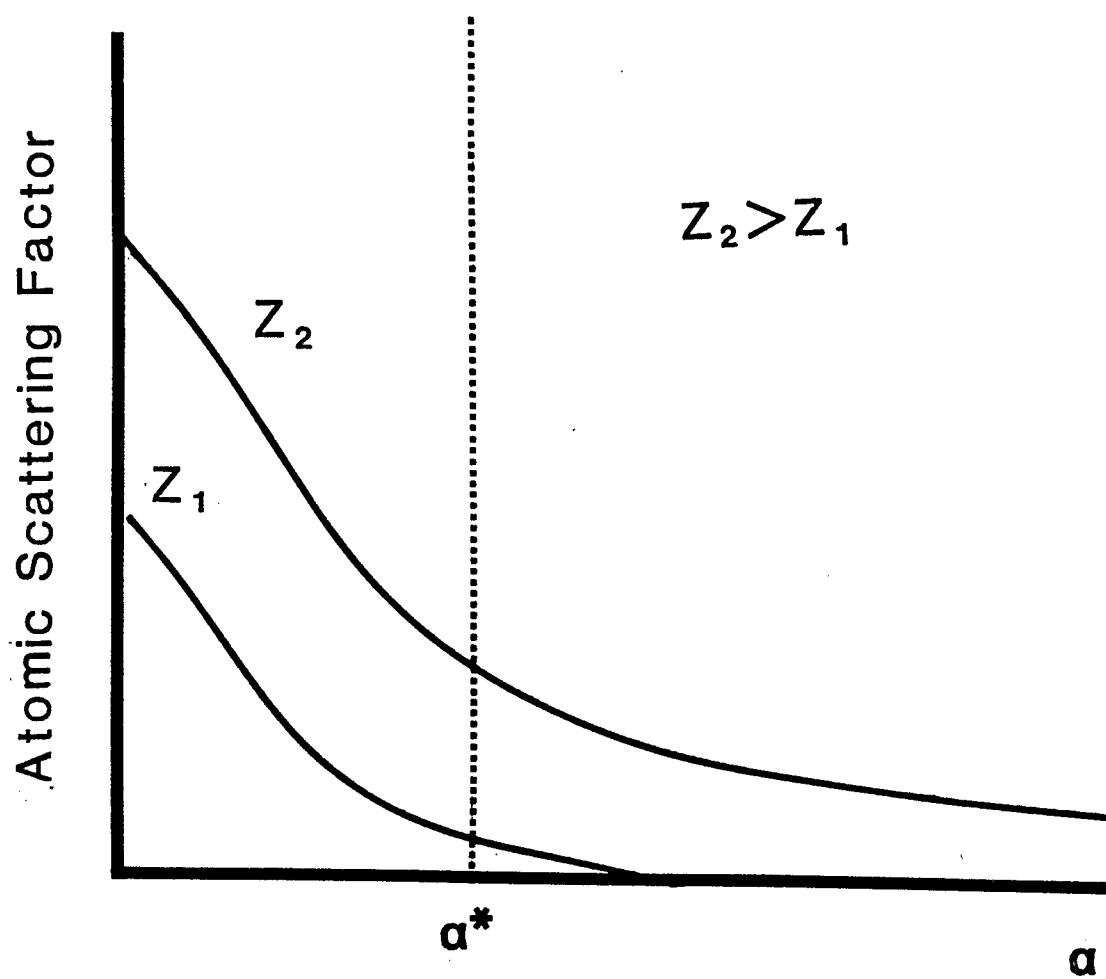Fig.1 - The iodine content in Iodopolystyrene as a function of reaction time.

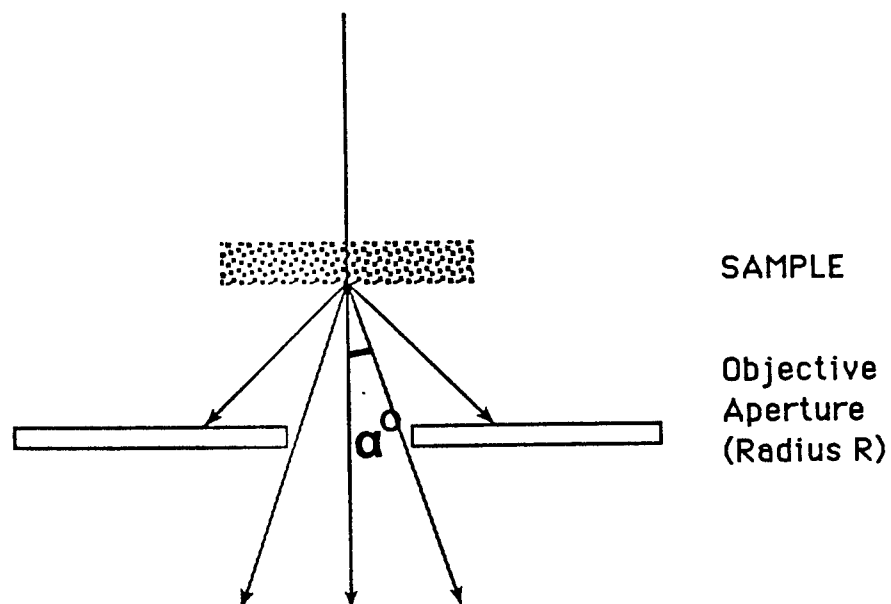Fig.2 - The atomic scattering factor as a function of the specific angle.

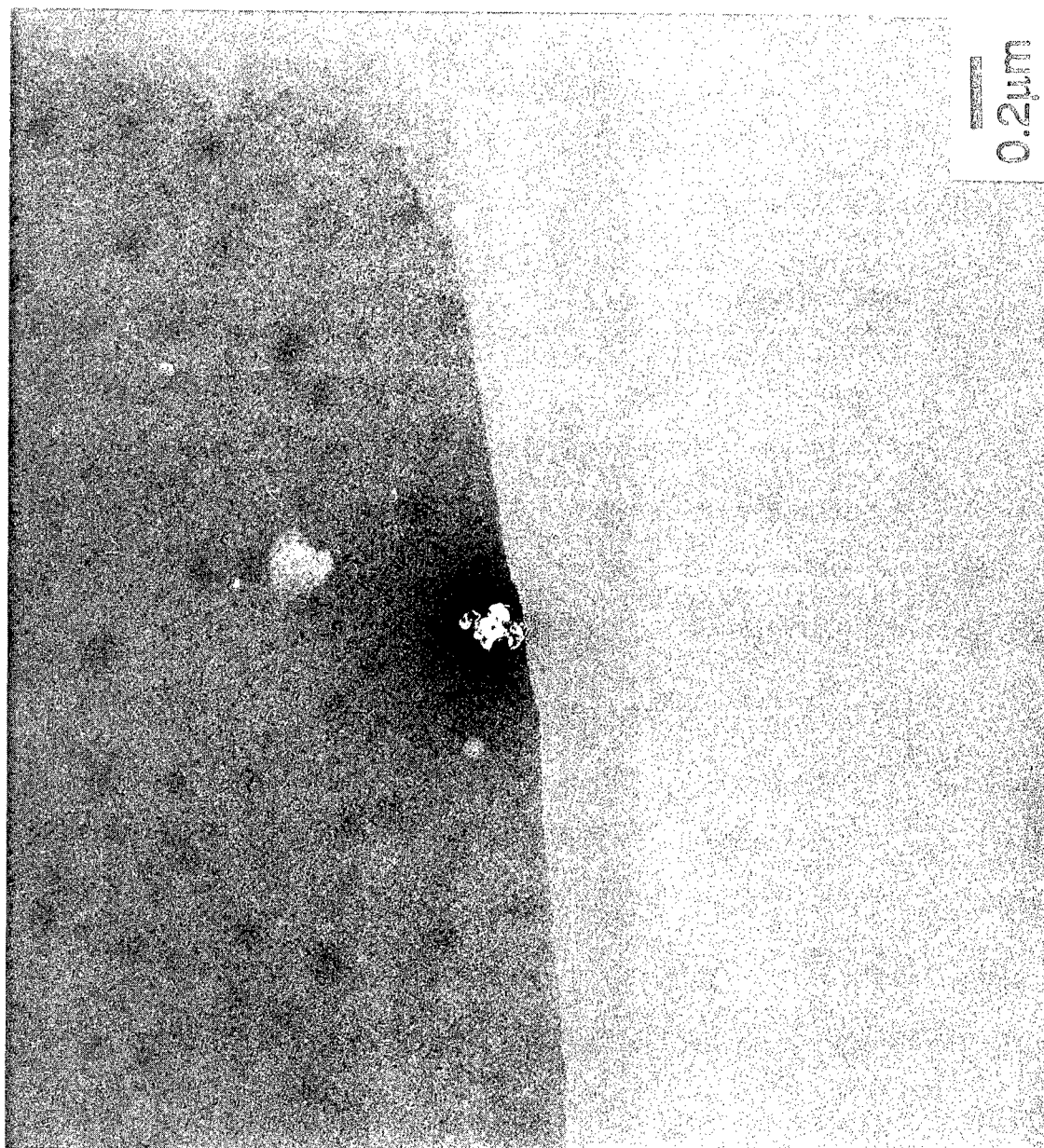Fig.3 - Relationship of critical angle α to the radius of the objective aperture R.

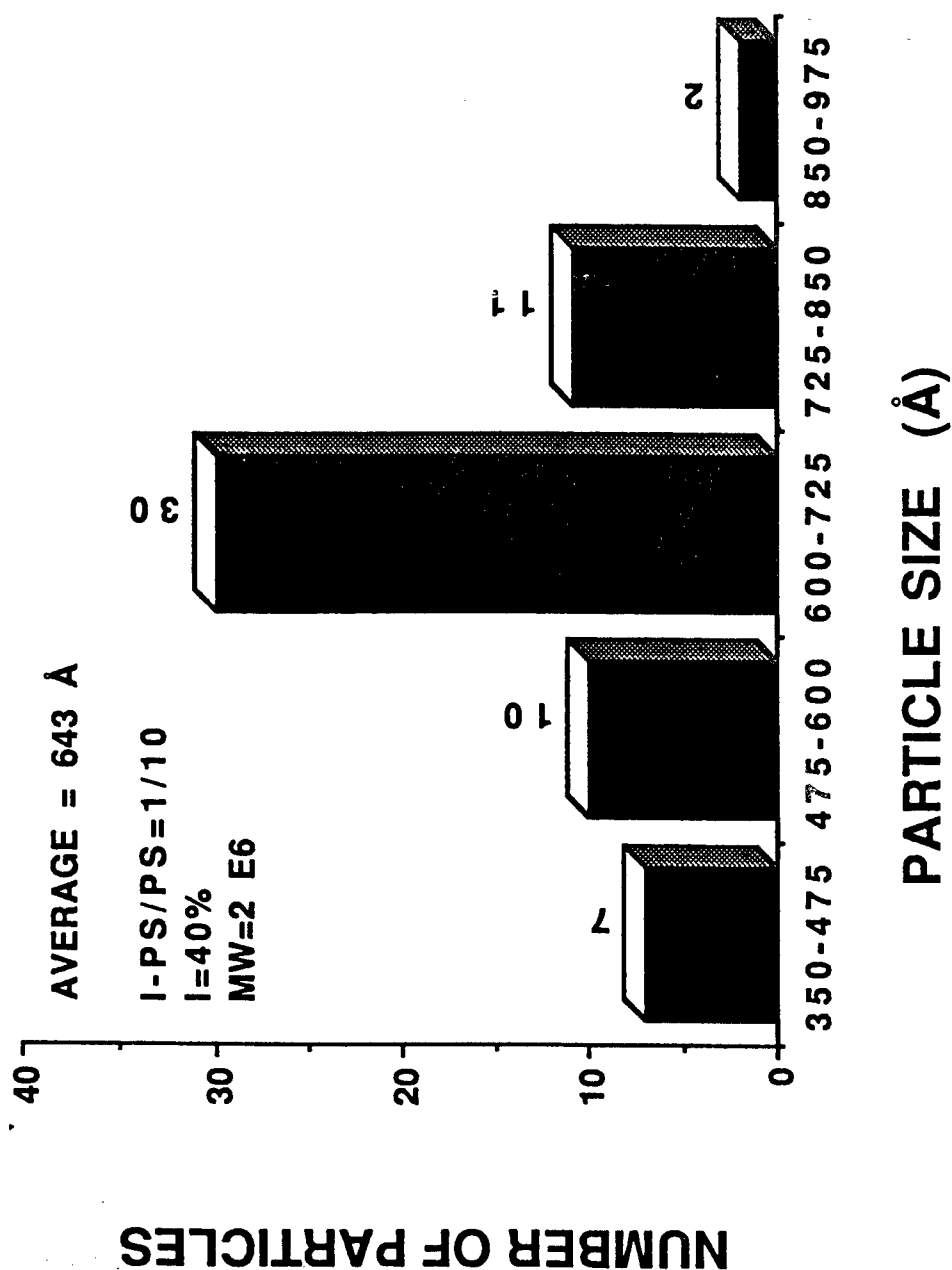Fig.4 - TEM photomicrograph for Iodopolystyrene to polystyrene ratio of 1:10.
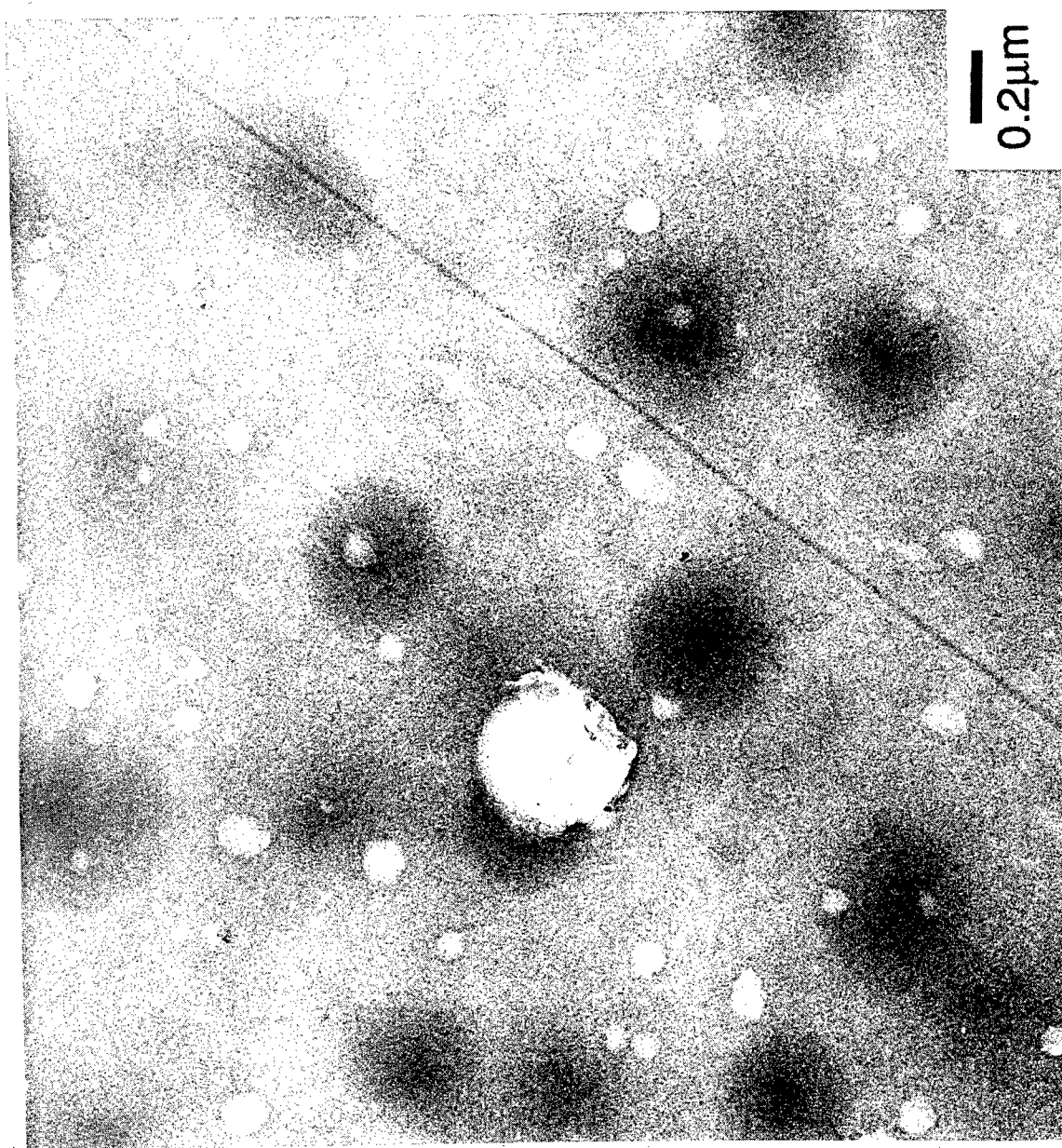
Fig.5 - Particle size analysis of ratio 1:10 (Fig.4).

Fig.6 - TEM photomicrograph for Iodopolystyrene to polystyrene ratio of 1:1.

Fuzzy Modeling of Shaped Charges Against Modern Armor (U)

Aivars Celmiņš, Dr.
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

Abstract.

The penetration capability of a shaped-charge jet is expressed by an integral over jet elements. The advantage of this formulation is that the effects of jet/armor interaction can be modeled individually for each jet element. That modeling is accomplished by a phenomenological interaction model, and the insufficient knowledge about details of the interaction is represented by defining an appropriate fuzziness of the model function.

1. Introduction.

A shaped charge consists of a metallic liner in a form that approximates a cone with its open base facing the target, and with explosive placed on the outside of the cone and behind it. The explosive is detonated at the rear end of the charge and, as the detonation front progresses toward the cone and target, the liner collapses and forms a jet. The jet has typically a diameter of the order of a millimetre and a tip velocity of the order of 7 km/s. Due to its high velocity, the jet is capable of penetrating steel armor to the order of several decimetres. The perforation depends on many factors one of which is the distance of the cone from the target. According to hydrodynamic theory, maximum penetration is achieved if that distance is of the order of several charge diameters. Experiments show that at larger distances the penetration depth decreases.

To counter the threat from shaped charges, a homogeneous steel armor may be supplemented with more or less complicated arrangements of armor plates made of different materials. Particularly effective are arrangements that include flying armor plates. In this paper, we consider the interaction of a shaped-charge jet with such a plate, and the effect of the collision on the jet's penetration capabilities. To model the effects of an assembly of plates, one might appropriately combine the effects of single interactions.

In Section 2 we describe mathematical models for jet penetration and interaction with a plate. Since the models are largely phenomenological we use fuzzy

functions for the implementation of the models. Section 3 provides a short outline of the concepts of fuzzy numbers and fuzzy functions. Section 4 presents an example of results with the final model.

2. Model Formulation.

2.1. Shaped Charge Penetration Model.

The particle velocities in a shaped-charge jet increase linearly with the distance of the particle from the charge. The point where the particle velocity equals zero is located within the liner cone and is called the *virtual center* of the jet. The positive velocity gradient within the jet causes the jet length to increase with time, and its diameter to decrease. When a critical diameter is reached then the continuous jet breaks up into small segments. The segments easily become distorted and disoriented and the penetration capability of the segmented jet is less than that of a continuous jet or of a sequence of well aligned and formed segments. Hydrodynamic theory predicts that the depth of penetration by an impinging undisturbed jet segment is proportional to the length of the segment. Therefore, the optimal distance of a shaped charge is such that most of the jet elements arrive at the target (the bottom of the penetration crater) around the breakup time when the jet is stretched but not yet distorted. The breakup time and the deterioration of the jet segment train depend on such things as the jet material, quality of the cone and explosive, and charge motion. In particular a rotation of the charge reduces its penetration capability as does a transverse velocity of the charge. Typically, the penetration/standoff relation for a given charge is derived from experiments. Figure 1 shows such an experimental curve. Before breakup, the penetration increases, the experimental scatter is small and the penetration agrees with theory. After breakup, the rate of increase diminishes and eventually the penetration decreases to zero. (For a detailed discussion of the penetration properties of shaped charges, see Held[1]). The initial part of the curve after breakup has been successfully represented by phenomenological models, and the final portion of the curve usually is determined by some form of data fitting or engineering insight. We propose a new model formulation that conforms with theory, precisely reproduces the experimental data and suites our purposes.

Let Λ_B [m] be the *length of the jet at breakup time* t_B [s] and λ [m] be a *material length coordinate* of jet elements such that $\lambda=0$ at the virtual origin and $\lambda=\Lambda_B$ at the tip of the jet. The speed of each jet element is

$$v(\lambda) = \frac{\lambda}{\Lambda_B} V_B \ ,$$
(2.1)

where V_B [m/s] is the speed of the tip. It is generally assumed that the penetration capability of a jet element is negligible if its speed is below a *cutoff speed* V_C. The corresponding cutoff λ-coordinate is
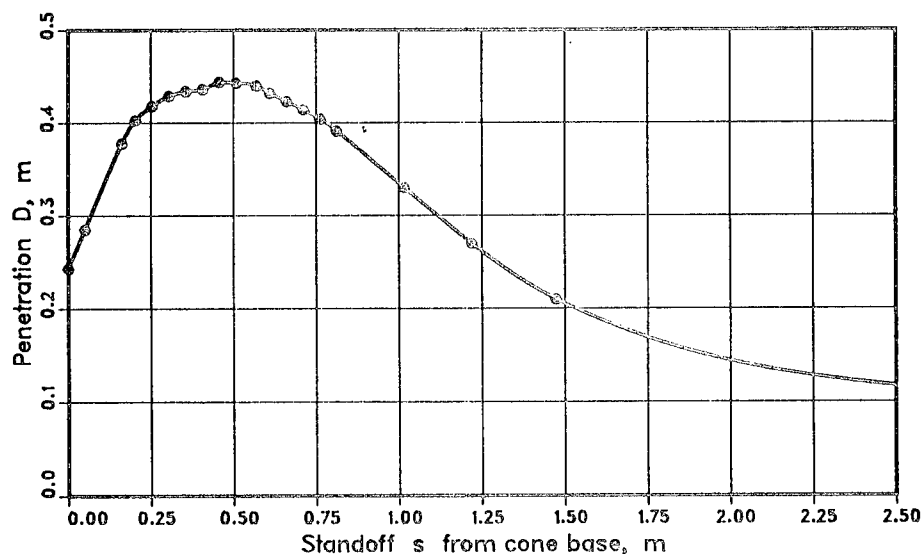
Figure 1. Penetration of solid armor.

$$\Lambda_C = \Lambda_B \; V_C \,/\; V_B \quad . \tag{2.2}$$

Let ρ_j [kg/m³] be the density of the jet material, ρ_a [kg/m³] be the density of the armor material and $\gamma = \sqrt{\rho_j / \rho_a}$. Let s [m] be the *virtual standoff*, i.e., the distance between the virtual center and the armor. The theoretical assumption that the crater depth increases proportionally to the length of the arriving element yields the following formula for the total depth D [m] of penetration by a continuous jet:

$$D(s) = \int\limits_{\Lambda_C}^{\Lambda_B} g(s,\lambda) \; d\lambda \quad , \tag{2.3}$$

where

$$g(s,\lambda) = \gamma \, \frac{s}{\Lambda_B} \, \left[\frac{\Lambda_B}{\lambda} \right]^{\gamma+1} \tag{2.4}$$

The dimensionless jet-element *effectiveness function* $g(s,\lambda)$ expresses the penetration capability of the jet element λ for given standoff s. One notices that the penetration capability of elements with smaller λ is greater. This is so because those elements arrive at the target later and therefore in a more stretched condition. One can show that a jet element λ arrives at the target in a continuous state if $g(s,\lambda) \leq \gamma$. Therefore, for given s, only elements with a λ-coordinate larger than

283

$$\Lambda_S(s) = \Lambda_B \left[\frac{s}{\Lambda_B}\right]^{1/(1+\gamma)} \tag{2.5}$$

arrive in continuous state. If after breakup the segments would not distort and misalign then the penetration capability of each element would remain constant and the corresponding penetration effectiveness function g would be

$$g_{theor}(s,\lambda) = \begin{cases} \gamma \cdot (s/\Lambda_B) \cdot (\Lambda_B/\lambda)^{\gamma+1} & \text{if } \lambda > \Lambda_S , \\ \gamma & \text{if } \lambda \le \Lambda_S . \end{cases} \tag{2.6}$$

Because an integration over this function predicts a too large penetration after breakup, we replace g_{theor} by the following function

$$g_{model}(s,\lambda) = \begin{cases} \gamma \cdot (s/\Lambda_B) \cdot (\Lambda_B/\lambda)^{\gamma+1} & \text{if } \lambda > \Lambda_S , \\ \gamma \cdot \left[\dfrac{\lambda + A \cdot \Lambda_S(s)}{(1+A) \cdot \Lambda_S(s)}\right]^n & \text{if } \lambda \le \Lambda_S , \end{cases} \tag{2.7}$$

with proper constants A and n. From inspection of typical penetration curves we have chosen $n=40$ and $A=19$. To obtain the observed $D(s)$ by integration over g_{model} we make the lower limit of the integral (2.3) a function $\Lambda_C(s)$ of the standoff. Hence, in our model we characterize the penetration capability of a shaped charge by using a cutoff function $\Lambda_C(s)$ instead of using the penetration function $D(s)$. The penetration function can be recovered by the integration (2.3) where the integrand is the function $g_{model}(s,\lambda)$ and the lower limit is the function $\Lambda_C(s)$. Figure 2 shows an example of the contour lines of the integrand g_{model} and the function $\Lambda_C(s)$ for the same warhead as in Figure 1. The tip of the jet has the coordinate Λ_B. For a given standoff, jet elements between Λ_B and the curve $g=\gamma$ arrive at the target in continuous state. Elements above the curve $g=\gamma$ arrive in segmented state. Jet elements above the curve $\Lambda_C(s)$ do not contribute to the penetration.

2.2. Models of Jet Interaction with a Plate.

We model the interaction between the jet and a plate by considering small segments $\Delta\lambda$ of the jet and treating each segment independently. A collision has the following effects:

(a) The plate is damaged, in particular, a slit might be cut or a hole might be pierced.

(b) The segment is deflected.

(c) The segment is eroded and otherwise damaged, e.g., by partial gasification so that its penetration capability is impaired.

The damage of the plate and the deflection of the segment are conceptually simple processes that are largely governed by laws of collision mechanics. The implementation of these laws by a computer program is more involved, because a number of special cases must be distinguished. Therefore, we present only a general outline of the interaction model in the case where a continuous jet cuts a slit in a flying plate.
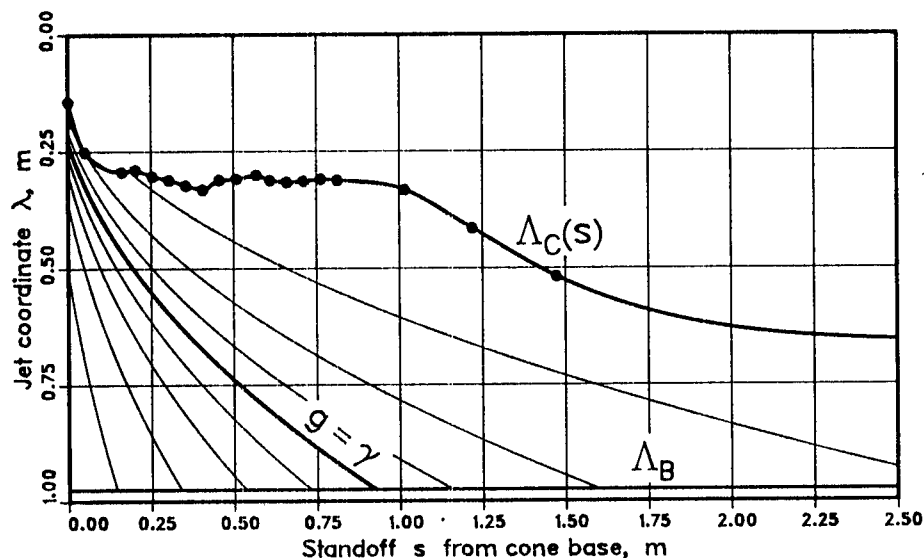
Figure 2. Penetration effectiveness function.

We chose a coordinate system with the x,y-plane parallel to the plate and specify the plate's position by the z-coordinate $z_p(t)$. Let the position of the jet element λ be given by the vector

$$\overrightarrow{X}_j(\lambda,t) = \begin{pmatrix} x_j(\lambda,t) \\ y_j(\lambda,t) \\ z_j(\lambda,t) \end{pmatrix} . \tag{2.8}$$

Let $t_k(\lambda)$ be the time of collision between the element and the plate. It is a solution of the equation

$$z_j(\lambda,\ t_k) = z_p(t_k) \ . \tag{2.9}$$

The x- and y-coordinates of the collision point are

$$x_k(\lambda) = x_j(\ \lambda,\ t_k(\lambda)\ ) \quad \text{and} \quad y_k(\lambda) = y_j(\ \lambda,\ t_k(\lambda)\ ) \ . \tag{2.10}$$

The results of the collision depend on the amount of masses involved. The mass of the jet segment is

$$\Delta m_j = \rho_j \cdot \pi \cdot r_j^2(\lambda) \cdot \Delta\lambda \ , \tag{2.11}$$

where $r_j(\lambda)$ [m] is the radius of the jet at breakup time. Let the plate have the thickness h_p [m], the density ρ_p [kg/m^3], and let the width of the slit be χ times the jet diameter. Let $r_k(\lambda)$ [m] be the radius of the jet at collision time. Then the amount of plate mass that is involved in the collision is

285

CELMIŅŠ

$$\Delta m_p = \rho_p \cdot h_p \cdot 2\, r_k \cdot \chi \cdot \Delta\lambda \cdot \left(\left[\frac{dx_k}{d\lambda} \right]^2 + \left[\frac{dy_k}{d\lambda} \right]^2 \frac{1}{\cos^2\omega} \right)^{1/2} , \qquad (2.12)$$

where the x,z-plane is assumed to contain the jet segment and ω is the angle between the direction of the jet and a normal to the plate (the *obliquity angle*). Let $\mu(\lambda) = \Delta m_p / \Delta m_j$ be the ratio of masses involved, $\vec{V}_j(\lambda)$ be the velocity of the jet element before collision and $\vec{V}_p$ be the plate's velocity. The velocity of the jet element after the collision is

$$\vec{V}_{j1}(\lambda) = \frac{1 - k \cdot \mu(\lambda)}{1 + \mu(\lambda)} \vec{V}_j(\lambda) + \frac{(1 + k) \cdot \mu(\lambda)}{1 + \mu(\lambda)} \vec{V}_p , \qquad (2.13)$$

where k is the collision coefficient. Eqs. (2.9) through (2.13) determine the initial location and velocity of each element after the collision. This information is used to compute the impact times and locations of the jet on other parts of the target. To apply these formulas one has to chose the slit size coefficient χ and the collision coefficient k. Both are phenomenological parameters with restricted ranges. The model also requires the determination of derivatives, see eq. (2.12).

Other entirely phenomenological models are used for the first impact of the jet on an undamaged part of the plate. If the plate is thick ($h_p > 2r_k$), then the area of the initial hole is approximately proportional to the jet footprint area times a function of a dimensionless "ballistic damage indicator" (Celmiņš[2]) which is defined as the kinetic energy density of the jet divided by the yield stress of the plate's material. Within the range of velocities that is typical for a shaped-charge jet, a linear function is appropriate and the crater radius r_p [m] can be estimated by

$$r_p(\lambda) = r_k(\lambda) \cdot \left[\beta \cdot \frac{\rho_j \cdot v^2(\lambda)}{2\, \sigma_p} \right]^{1/2} , \qquad (2.14)$$

where β is a model parameter and σ_p [Pa] is the yield stress of the plate. If the plate is very thin then a better approximation is $r_p = \chi\, r_k$. Therefore, the complete model includes a transition from thin to thick plate models.

The hydrodynamic erosion formula for the tip of the jet is

$$\Delta\lambda_E = \frac{t_B - t_0}{t_k(\lambda) - t_0} \cdot \left[\frac{\rho_p}{\rho_j} \right]^{1/2} \cdot \frac{h_p}{\cos\omega} , \qquad (2.15)$$

where t_0 is the ignition time. Analogous to the hole size model, the erosion model includes a transition from near zero erosion for thin plates to the hydrodynamic erosion formula (2.15) for thick plates.

The exercise of the interaction models provides a description of the jet after the collision. In general, after collision, the jet is not straight and has a transverse velocity component, two properties that impair its penetration capability.

2.3. Disturbed Jet Penetration Model.

To express the penetration of solid armor by a disturbed jet we modify the penetration integral (2.3) as follows.

$$D_m(s) = \int_{\Lambda_c(s)}^{\Lambda_B - \Delta\lambda_E} g_{model}\big(s_a(s,\lambda),\lambda \big) \cdot \eta \cdot I(\lambda) \cdot d\lambda \ . \tag{2.16}$$

The modifications are:

(1) The upper integration limit is $\Lambda_B - \Delta\lambda_E$ instead of Λ_B to model the erosion of the tip.

(2) The argument s of the effectiveness function g_{model} is replaced by an *effective virtual standoff*

$$s_a(s,\lambda) = \lambda \cdot \frac{t_a(s,\lambda) - t_0}{t_B - t_0} \ , \tag{2.17}$$

where $t_a(s,\lambda)$ is the arrival time of the jet element at the armor. The effective standoff is the distance of the point of impact from the virtual center. If the jet is not disturbed then $s_a = s$. If the jet element is deflected then the distance can be different because the deflection changes the impact location.

(3) The effectiveness reduction factor η models the loss of penetration effectiveness of a jet element due to suffered collisions, e.g., by partial gasification or other unknown effects. The factor might depend on λ, t_k, or on other pertinent arguments, e.g., on the number of collisions suffered by the element before it arrives at the target.

(4) The relative intersection area $I(\lambda)$ models the effect of transverse dispersion on jet performance. The idea is to integrate only over those parts of the jet that are not deflected out of the main crater in the target. $I(\lambda)$ is the area of intersection of the jet footprint with the main target hole entry, divided by the area of the jet footprint. For this calculation, the "main target hole entry" is the entry to the deepest crater. Usually it is at the location where the first jet element arrives at the target.

3. Fuzzy Models.

3.1. Fuzzy Numbers and Vectors.

Fuzzy numbers were introduced by Zadeh in 1965[3] as a means to represent imprecise information that is not probabilistic. Such information can be, for instance, intelligence data or a future concept of a weapon system. An often used representation of this type of information is by interval numbers that are a formalization of the information "a is known to be between a_1 and a_2". The corresponding interval number is the set of all those numbers x that are elements of the interval (a_1, a_2). To represent more detailed approximate information, such as in "a is about a_3, but certainly above a_1 and below a_2", one can assign to each x

between a_1 and a_2 a positive membership value $\mu(x)$ as a measure of the possibility that a assumes that particular value of x. This defines a set of numbers within (a_1,a_2) whereby the membership value $\mu(x)$ measures a degree of how much x belongs to the set. Sets that are defined in terms of a non-constant membership function are called *fuzzy sets*. Usually the membership is normalized by setting the maximum membership value equal to unity. Thus, in the given example one would set $\mu(a_3)=1$, assign smaller positive membership values to other x within the interval (a_1,a_2) and make $\mu(x)\equiv 0$ outside the interval. This fuzzy set of real numbers is called a *fuzzy number*. An interval number is the special case of a fuzzy number where the membership function $\mu(x)\equiv 1$ if x is within (a_1,a_2). If the length of the interval is reduced to zero then one has a regular (crisp) number. Fuzzy sets permit one to precisely define the meaning of such statements as "about two hours", "reduced functionality", "similar to a second degree function", etc. and allows one to operate with the imprecise information in a computing environment. We now present a short outline of some formal definitions related to fuzzy numbers and vectors that are used in the present model.

A fuzzy number $\tilde{a}$ is a set of numbers $w \in {I\!R}^1$ whereby each element of the set is assigned a membership value $\mu_a(w)$ between zero and one. All fuzzy numbers considered in this paper are assumed to have triangular membership functions

$$\mu_a(w) = 1 - \min\left\{\, 1,\ |w-a|\,/\,s_a \,\right\}\ , \tag{3.1}$$

where s_a is a positive constant. Elements of the fuzzy set $\tilde{a}$ are all such numbers w for which $\mu_a(w) > 0$. With this definition, a fuzzy number is completely specified by two real numbers, a and s_a:

$$\tilde{a} = \left\{\, a,\ s_a \,\right\}\ . \tag{3.2}$$

We call a the *apex* of $\tilde{a}$ and s_a the *spread* of $\tilde{a}$.

A generalization to fuzzy vectors is as follows[2]. Let $x \in {I\!R}^n$, $A \in {I\!R}^n$ and let P_A be a positive definite $(n \times n)$-matrix. We associate with A the following elliptic norm for the distance between x and A:

$$\|x-A\|_A = \left[\, (x-A)^T\, P_A^{-1}\, (x-A) \,\right]^{1/2}\ , \tag{3.3}$$

and define a conical membership function by

$$\mu_A(x) = 1 - \min\left\{\, 1,\ \|x-A\|_A \,\right\}\ . \tag{3.4}$$

The fuzzy vector $\tilde{A}$ is the set of all vectors x with the property $\mu_A(x) > 0$. A fuzzy vector $\tilde{A}$ with a conical membership function is completely specified by the elements of A and P_A:

$$\tilde{A} = \left\{\, A,\ P_A \,\right\}\ . \tag{3.5}$$

We call A the apex of $\tilde{A}$ and P_A the *panderance matrix* of $\tilde{A}$. Each component of $\tilde{A}$ has a triangular membership function, the projection of the cone (3.4) onto the corresponding axis. Its spread is the square root of the corresponding diagonal

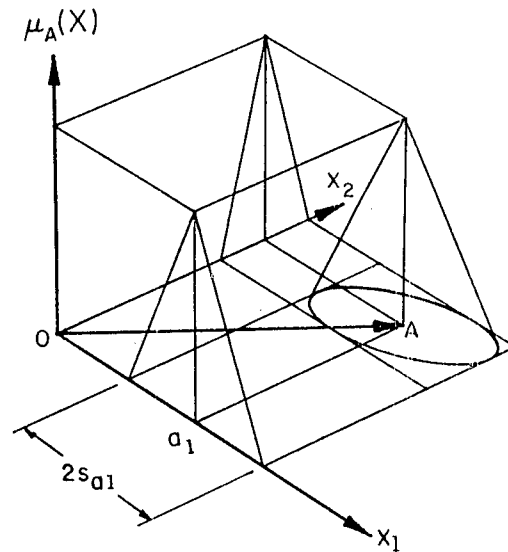element of P_A. A schematic view of the membership function in $I\!\!R^2$ is shown in Figure 3.

Figure 3. Conical membership function in $I\!\!R^2$.

3.2. Fuzzy Functions.

A fuzzy function is a fuzzy set of functions, that is, a set of functions whereby each element of the set is assigned a membership value. In this paper, we specify the set in terms of a fuzzy parameter vector as follows[2].

Let $x\in I\!\!R^n$, $t\in I\!\!R^p$ and $F(x,t)$ be a r-dimensional function. We consider t as a parameter which defines a set of functions of x. Let P_T be a positive definite $(p\times p)$-matrix, $T\in I\!\!R^p$ and

$$\widetilde{T} = \{\ T,\ P_T\ \} \tag{3.6}$$

be a fuzzy vector in $I\!\!R^p$. We assign to each element of the set of functions the same membership value as the corresponding parameter:

$$\mu_F\big(\ F(x,t)\ \big) = \mu_T(t)\ . \tag{3.7}$$

This defines a fuzzy set of functions

$$\widetilde{F}(x) = F(x,\widetilde{T})\ . \tag{3.8}$$

A crisp equation $F(x,t){=}0$ defines for fixed t a $(n-r)$-dimensional crisp surface in $I\!\!R^n$. The corresponding fuzzy equation

$$\widetilde{F}(x) = F(x,\widetilde{T}) = 0 \tag{3.9}$$

defines a fuzzy surface, that is a set of surfaces with corresponding membership

values. The surfaces of this set fill a region in $I\!\!R^n$ and we define that region as the fuzzy solution $\widetilde{X}_F$ of the fuzzy equation $\widetilde{F}(x)=0$. To completely describe the solution, we assign specific positive membership values for those x that are within the region. A given x usually is either not a solution of any of the equations (3.9) in which case we assign to it a zero membership value, or it is a solution for a finite range of parameter values t in which case we assign to x the largest membership value of that range:

$$\mu_{XF}(x) = \max_{t:F(x,t)=0} \mu_F\big(F(x,t) \big) \ . \tag{3.10}$$

This defines $\widetilde{X}_F$ as a fuzzy set of points in $I\!\!R^n$, but that set does not necessarily have a conical membership function and we shall not call it a fuzzy vector. The surfaces $\mu_A(x)=constant$ are hyperellipsoids (see Figure 3) whereas the surfaces $\mu_{XF}(x)=constant$ are in general approximately cylinders with rulings parallel to the apex surface $F(x,T)=0$.

To have a more convenient definition of μ_{XF}, analogous to (3.4), we define a norm for the separation of an x from the fuzzy solution $\widetilde{X}_F$ by

$$\|x-\widetilde{X}_F\|_{XF} = \min_{t:F(x,t)=0} \|t-T\|_T \ . \tag{3.11}$$

To be definite, we also assign an infinite value to $\|x-\widetilde{X}_F\|_{XF}$ if $F(x,t)=0$ has no solution for the given x. With this definition, eq (3.10) is identical to

$$\mu_{XF}(x) = 1 - \min\big\{ 1, \|x-\widetilde{X}_F\|_{XF} \big\} \ . \tag{3.12}$$

An approximate formula for (3.11) is[2]

$$\|x-\widetilde{X}\|_{XF} \approx \big[F^T (F_t P_T F_t^T)^{-1} F \big]^{1/2} \ , \tag{3.13}$$

where $F_t=\partial F/\partial t$ is the Jacobian matrix of F, and F and F_t are evaluated at (x,T). If $F_t P_T F_t^T$ is semi-definite then the Moore-Penrose generalized inverse is used in (3.13). The formula is exact is F is a linear function of t.

3.3. Fuzzy Models.

A fuzzy predictive model typically is used in the form of a function

$$\widetilde{Z}(x) = F(x,\widetilde{T}) \ , \tag{3.14}$$

where x is an explanatory variable and $\widetilde{T}$ is a fuzzy parameter of the model function. What one needs in this case is a formula for the propagation of the fuzziness of $\widetilde{T}$. It is provided by the *extension principle*[4] which states that if $z=G(t)$ and the argument t is a fuzzy $\widetilde{T}$, then also z is a fuzzy $\widetilde{Z}$ and the membership function of $\widetilde{Z}$ is

$$\mu_Z(z) = \max_{t:z=G(t)} \mu_T(t) \ . \tag{3.15}$$

In our case, $G(t)=F(x,t)$ and, if t is a fuzzy vector $\widetilde{T}$ with a conical membership function, then to a first approximation also z is a fuzzy vector $\widetilde{Z}$ with the apex $Z=F(x,T)$ and the panderance matrix[2]

290

$$P_Z \approx F_t P_T F_t^T \qquad (3.16)$$

The panderance propagation formula (3.16) is exact and consistent with the extension principle if F is a linear function of t. If the panderance matrix P_Z is only semi-definite then the membership function of $\tilde{Z}$ is a degenerated cone.

If a mathematical model is a crisp function, then all algebraically equivalent formulations of the model function (3.14) produce the same result. If the model function is fuzzy, then only the apex of the prediction is invariant to algebraic manipulations whereas the fuzziness of the result depends on the formulation, as can be seen from eq. (3.16).

4. Example.

4.1. Implementation of the Interaction Model.

The interaction model is in essence determined by the following set of parameters.

β — crater size model coefficient, eq. (2.14),
η — penetration effectiveness reduction factor, eq. (2.16),
χ — plate slit size coefficient, eq. (2.12),
k — collision coefficient, eq. (2.13).

The values of the parameters were established by comparing the predicted $D_m(s)$, eq. (2.16), with observed $D(s)$ for a number of cases involving modern armor. In addition, data were used from experiments involving shaped charges with transverse velocities, for instance from Golesworthy 1983[5]. The modified integral (2.16) models the effects of transverse velocity via the relative intersection area function $I(\lambda)$. Transverse velocity experiments allow one, therefore, to estimate the crater size parameter β independently of the other parameters.

The model has been implemented at the Ballistic Research Laboratory as a software package that computes the predicted penetration for given weapon and target geometry. An indication of the complexity of the modeled problem is that one needs, in addition to the empirical penetration function $D(s)$, over 30 parameters to completely describe a typical weapon/armor combination. Consequently, it is practically impossible to obtain an exhaustive matrix of experiments that covers all relevant cases, and estimates of the model parameters must be based partly on experience and judgement, a typical situation for the application of fuzzy set representation. Therefore, all model parameters are assumed to be fuzzy numbers with appropriate spreads, that is, the penetration model is fuzzy in the sense of eq. (3.14).

4.2. Illustrative Examples.

To illustrate the working of the model, we compare in Figure 4 the integrand $g_{model}(s,\lambda)$ with the modified integrand in eq. (2.16). The heavy top curve in Figure 4 is the function g_{model} for $s=0.50$ m, that is, a vertical slice through the surface
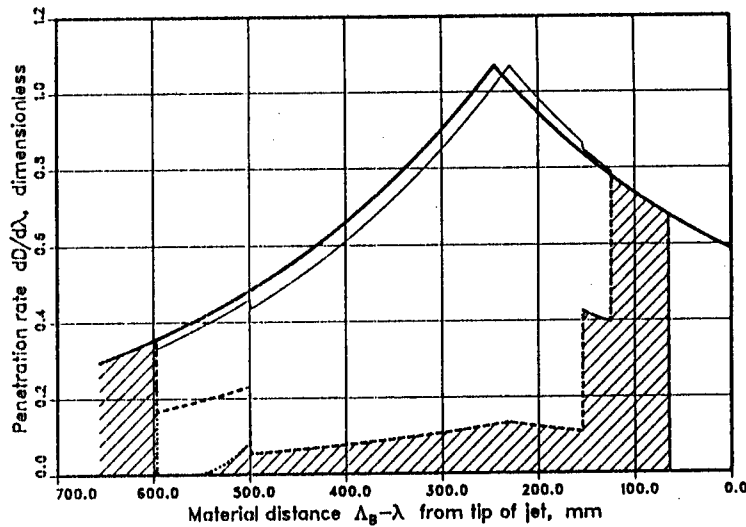
291

Figure 4. Penetration rate for constant standoff.

pictured in Figure 2. The total penetration by an undisturbed jet equals the area below the curve. The function g_{model} has a maximum value that equals γ. Elements to the right from the maximum arrive at the target in continuous form. Elements to the left from the maximum arrive in segmented form, and later arriving elements contribute less to the penetration, as discussed above. The modifications of the integrand due to interactions with plates are indicated by the several other curves that represent reduced integral values due to the various effects discussed above. The modified integrand in eq. (2.16) is not continuous, because different effects are acting on different parts of the curve, e.g., the tail end of the jet is not affected after its front end has cut completely through a plate with finite width. The predicted residual penetration D_m in the main armor is the shaded area. It is the depth of penetration of the jet into a main solid armor after it has penetrated an assembly of plates.

The curves shown in Figure 4 are those of a crisp model. If the model is fuzzy, then the shaded area has a fuzzy upper boundary. The fuzziness of the boundary can be estimated from the fuzziness of the model parameters using the panderance propagation formula (3.16). The corresponding integral (the area under the curve) is a fuzzy number. Figure 5 shows examples of the main armor penetration versus standoff in cases of no interaction and interaction with a plate assembly. The top curve in Figure 5 is the same as shown in Figure 1. The width of the vertically hatched area corresponds to 90% probability of penetration to a depth within the spread according to experimentally determined scatter. The bottom curve shows the predicted penetration $\hat{D}_m(s)$ by a disturbed jet. The curve is fuzzy by definition and its spread that stems from the fuzziness of the model parameters is indicated by the diagonally hatched area.
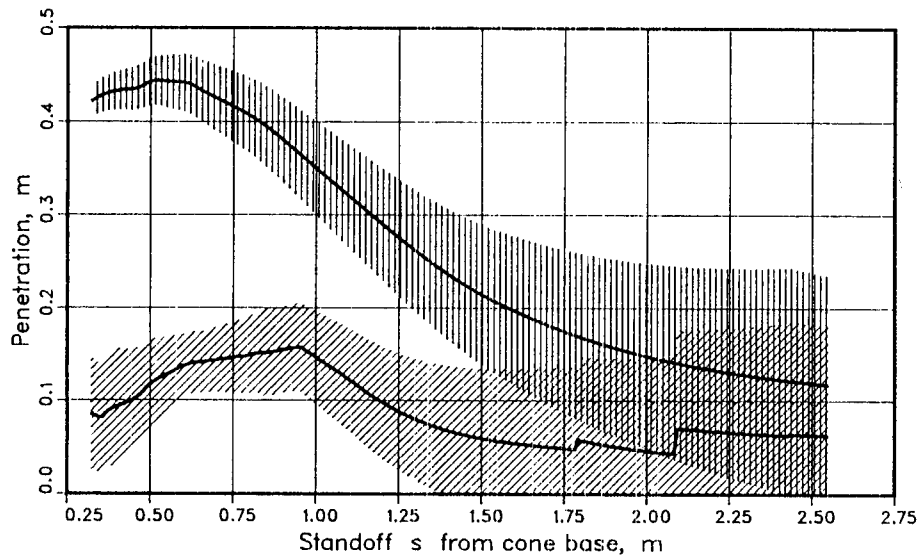
Figure 5. Effects of modern armor.

In Figure 6 we compare the predicted residual penetration with results from 94 experiments with different types of armor.
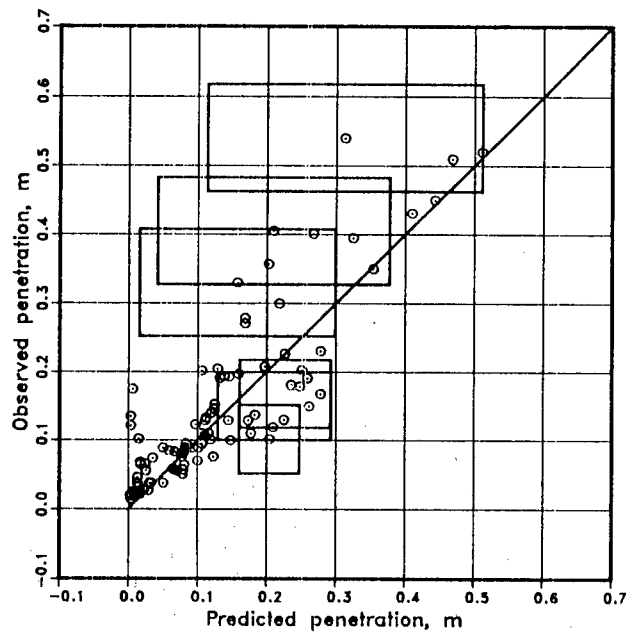

Figure 6. Predicted and observed penetrations.

To illustrate the accuracies of the observations and predictions we have plotted rectangles around some data points with large differences between predictions and observations. The widths of the rectangles equal the spread of the fuzzy predictions and the heights of the rectangles are estimated 90% scatter intervals of observations. The figure indicates that for the given data set the model accuracy is of the same order as the experimental accuracy.

5. Summary and Conclusions.

A mathematical model has been developed for the interaction of a shaped-charge jet with a flying armor plate. The model is combined with an integral representation of armor penetration to yield the effects of modern armor assemblies on the penetration effectiveness of shaped charges. Because exhaustive experimental testing is not feasible, parts of the model are based on estimates by judgement. To represent the vagueness of such estimates, fuzzy functions are used and the predicted penetration estimates are fuzzy numbers.

Comparisons between predicted and observed penetrations show that the overall accuracy of the prediction is comparable with experimental accuracies. An advantage of the present model is its capability to estimate the effect on warhead performance of the synergism between transverse velocity and modern armor.

References.

[1] M. Held, "Characterizing Shaped Charge Performance by Stand-Off Behavior', *Proceedings of the Seventh International Symposium on Ballistics*, The Hague, The Netherlands (1983), pp. 331-333.

[2] A. K. Celmiņš, "Multidimensional Least-Squares Fitting of Fuzzy Models", *Mathematical Modelling* 9 (1987), pp. 669-690.

[3] L. A. Zadeh, "Fuzzy Sets", *Information and Control* 8 (1965), pp. 338-353.

[4] L. A. Zadeh, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning", Part I, *Information Sciences* 8 (1975), pp. 199-249.

[5] R. C. Golesworthy, "The Effect of Transverse Velocity on Shaped Charge Performance", *Proceedings of the Seventh International Symposium on Ballistics*, The Hague, The Netherlands (1983), pp. 257-260.

Dynamical Behavior of Solid Explosives

*Cary F. Chabalowski, Dr.
George F. Adams, Dr.
Betsy M. Rice, Dr.
Donald L. Thompson[a], Dr.
Ballistic Research Laboratory, SLCBR-IB-I,
Aberdeen Proving Ground, MD 21005-5066
a. Department of Chemistry, Oklahoma State University,
Stillwater, OK 74078

I. INTRODUCTION: Energetic compounds containing nitro groups exhibit complex chemical behavior, and resolution of the mechanistic details of the reactions of these compounds at a molecular level is an experimental and a theoretical challenge. The chemistry of the simplest organo-nitro compound, nitromethane, is not completely understood, despite several studies[1] done on the system. Even seemingly direct studies of the unimolecular decomposition of gas-phase nitromethane have indicated complexities in the mechanism of decomposition.[2,3]

The infrared multiphoton dissociation (IRMPD) experiments of Wodtke, Hintsa and Lee[2] indicate that isomerization of nitromethane to methyl nitrite competes with scission of the CN bond. Assuming an isomerization channel and using RRKM theory to fit the experimental results, Wodtke, et al., predict a 55.5 kcal/mol barrier to isomerization, which is slightly lower than the C-N bond dissociation energy (59.5 kcal/mol). They conclude that unimolecular decay of nitromethane can occur in either of two competing reactions:

$$CH_3NO_2 \rightarrow CH_3NO_2^\dagger \rightarrow CH_3 + NO_2 \qquad\qquad (I)$$

$$CH_3NO_2^\dagger \rightarrow CH_3ONO^\dagger \rightarrow CH_3O + NO \quad . \qquad\qquad (II)$$

There is a conflict between the experimental and theoretical value for the barrier to isomerization with the theoretical value estimated to be 76 kcal/mol.[4-6] It has been suggested that another transition state exists that is more of a van der Waals complex, with longer C-N and C-O bonds (approximately 4.0 Å).[7,8] Unfortunately, there are no reliable data describing the nature of such a transition state.

We present the results of a classical dynamics calculation of the unimolecular decomposition of nitromethane based on the reaction scheme suggested by Wodtke, et al.[2] Because the energy and the geometry of the transition state is not established, we have constructed three model potential energy surfaces to describe nitromethane, which differ primarily in the height of the isomerization barrier. The three barriers are chosen to be (1) 216.4 kcal/mol, (2) 55.1 kcal/mol, and (3) 47.6 kcal/mol. An interesting result of this study is that all three surfaces predict that unimolecular decay of CH_3NO_2 occurs not just through Reactions I and II, but also by the reaction

$$CH_3NO_2^\dagger \rightarrow CH_3ONO^\dagger \rightarrow CH_3 + NO_2 \quad . \tag{III}$$

II. POTENTIAL ENERGY SURFACE: The potential energy is:

$$V_{Total} = \Sigma \; V_i S_i \tag{1}$$

where i denotes all the reactants and decomposition products. The V_i describe the stable species i at equilibrium, and is made up of terms which describe stretching and bending motions. The S_i is a weighting function which is dependent on critical geometries of the species. This function increases or decreases the contribution of the V_i to the overall potential depending on the molecular geometry.

An advantage to using this form of potential is that the weighting functions can be monitored throughout a trajectory to characterize the system at any time. This provides an easy way to determine a reaction mechanism in a complex system without having to rely on plots of individual internal coordinates as functions of time. Another advantage to using this type of potential is that it is "modular"; the different V_i developed can be used in other dynamics studies which use a potential energy surface of the form in Eq. (1). If other channels need to be added, all that is necessary is that the new $V_i S_i$ be added to Eq. (1), and the existing weighting functions modified accordingly.

Expanding Eq. (1), the total potential energy function to describe the nitromethane/methyl nitrite system is:

$$V_{Total} = V_{CH_3NO_2} S_{CH_3NO_2} + V_{CH_3ONO'} S_{CH_3ONO'} + V_{CH_3O'NO} S_{CH_3O'NO}$$

$$+ V_{CH_3+NO_2} S_{CH_3+NO_2} + V_{CH_3O+NO'} S_{CH_3O+NO'} + V_{CH_3O'+NO} S_{CH_3O'+NO} . \tag{2}$$

296

Note that there are two terms to describe the methyl nitrite species, and two terms to describe the product fragments CH_3O and NO. The terms differ only in the labeling of the oxygens, which are labelled O and O'. These are chemically equivalent species and in reality cannot be distinguished. A dynamics study must allow unbiased formation of methyl nitrite and the product fragments with either O or O' bound to the carbon atom. Therefore, it is necessary, according to the formulation of the potential energy surface given in Eq. (1), to have terms which explicitly describe the species in which either O or O' is bound to the carbon atom. The parameters of the potential function are fit to the known spectroscopic, kinetic and thermodynamic data. Additionally, selected points calculated by <u>ab initio</u> methods are incorporated into the fit.

III. FEATURES OF THE POTENTIAL ENERGY SURFACE: Equilibrium geometries and fundamental vibrational frequencies of all stable species involved in the study have been reproduced by these potential energy functions. In addition to giving reasonable values for the equilibrium properties of all stable species of interest, the three surfaces correctly describe the thermochemistry of the reactions. Table 1 gives the relative energies of all reactant and product species of interest in this study. Additionally, important structural changes associated with the various bond ruptures or internal rotations of the species are described by the potential energy surfaces.

Table 1. Relative Energies (kcal/mol) of Species on the CH_3NO_2
Potential Energy Surface

Molecule	PES 1	PES 2	PES 3	Expt'l.
CH_3NO_2	0.0	0.0	0.0	0.0
cis-CH_3ONO	2.30	3.04	3.04	2.3[9]
trans-CH_3ONO	3.36	3.36	3.36	3.1[10]
CH_3O + NO	43.0	43.0	43.0	43.5[11]
CH_3 + NO_2	60.0	60.0	60.0	59.5[9,11]
Transition State	216.4	55.1	47.6	55.5[2]

Figures 1(a,b) are contour plots of the potential energy as a function of the CNO angle and the C–N bond distance for PES 1 and PES 3, respectively. For each grid point (i.e., C–N bond distance and CNO angle) in these plots, all other internal coordinates are allowed to relax to the equilibrium position. Regions which energetically describe the species CH_3NO_2, CH_3ONO and CH_3 + NO_2 are all shown on these plots. The nitromethane well is located in the lower right-hand corner of the figures; the minimum of this well is at R(C–N) = 1.479 Å and Θ_{CNO} = 117.1° and had a relative energy value of 0.0 kcal/mol. The methyl

nitrite well is located in the lower left-hand corner of the figures; the minimum of this well is at $R(C-N) = 2.3$ Å and $\Theta_{CNO} = 32°$ and has an energy value of 3.4 kcal/mol. The flat region which runs across the upper portion of the contour plots at a constant value of 60.0 kcal/mol corresponds to the CH_3+NO_2 region of the surfaces.

The transition state for isomerization of nitromethane to methyl nitrite has been located on each of the three potential energy surfaces. On PES 1 (Fig. 1a), the transition state is located at $R(C-N)$ = 1.567Å and $\Theta_{CNO} \approx 75°$, and has a value of 216.4 kcal/mol. In this study isomerization to methyl nitrite from nitromethane cannot occur on PES 1 via the transition state because all calculations were done at energies much lower than 216.4 kcal/mol. Isomerization can occur if a trajectory undergoes C-N bond scission and leaves the nitromethane well, passes through the flat region of the potential (the $CH_3 + NO_2$ region), and then moves inward to the methyl nitrite well. For this surface, isomerization is a two-step process: bond scission to $CH_3 + NO_2$ followed by recombination to methyl nitrite.

Because PES 2 and PES 3 have similar features except for the barrier heights, we choose to show only the contour plot for PES 3. The transition state on PES 2 has a value of 55.1 kcal/mol, the value suggested by Wodtke, et al.,[2] and is located at $R(C-N)=2.276$Å and $\Theta_{CNO}=62.2°$. The transition state on PES 3, shown in Fig. 1b, is located at $R(C-N)=2.196$Å and $\Theta_{CNO}=61.9°$, and has an energy of 47.6 kcal/mol, which is the value predicted from the MINDO/3 calculations of Dewar and Ritchie.[12] The dissociation-recombination mechanism can also occur on these surfaces, but the direct isomerization, i.e., concerted C-N bond scission and C-O bond formation via the transition state, is more probable.

IV. DETAILS OF THE CALCULATIONS: Ensembles of trajectories with initial energies in the range 64.6 to 161.4 kcal/mol above the zero-point energy of nitromethane (29.3 kcal/mol), were integrated using a variable-step size Adams-Moulton fourth-order predictor-corrector integrator.[13] Relative error tolerance was set at 10^{-7}. The initial conditions were selected using Metropolis Monte Carlo sampling.[14,15] Ensembles of 1000 trajectories at specific energies were run on each of the three potential energy surfaces described in Section II, which differ in their barrier heights to isomerization. A trajectory is integrated until the end tests are satisfied, or until the trajectory integration exceeds a maximum time of 30 ps. For all three surfaces, and for most energies, Reactions (I), (II), and (III) were observed.

V. RESULTS AND DISCUSSION: We have observed two mechanisms for isomerization in this system. The first mechanism is illustrated in a typical trajectory which produces the products CH_3O+NO on PES 1. The

mechanism of isomerization is dissociation to CH_3 + NO_2 followed by recombination, to form CH_3ONO. The total energy of the trajectory is 69.2 kcal/mol above the zero-point of nitromethane. Figures 2a(i-iv) are the weighting functions $S_{CH_3NO_2}$, $S_{CH_3+NO_2}$, $S_{CH_3O'NO}$, and $S_{CH_3O'+NO}$, respectively, vs. time. For the first 0.6 ps of the trajectory, $S_{CH_3NO_2}$ is unity, while all other S_i are zero. $S_{CH_3NO_2}$ sharply decreases to zero as $S_{CH_3+NO_2}$ concurrently rises to one, indicating that the total potential energy is now characterized mainly by the terms which describe CH_3 + NO_2. After approximately 0.1 ps, however, $S_{CH_3+NO_2}$ drops to zero as $S_{CH_3O'NO}$ rises to unity indicating formation of methyl nitrite where O' is bound to the carbon atom. After 1.75 ps, $S_{CH_3+NO_2}$ drops to zero while $S_{CH_3O'+NO}$ rises to one as CH_3O + NO is formed, giving a reaction mechanism of:

$$CH_3NO_2 \rightarrow CH_3 + NO_2 \rightarrow CH_3ONO \rightarrow CH_3O + NO.$$
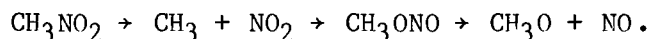
This interpretation of the reaction is confirmed by Figs. 2a(v-xi) which show various internal coordinates versus time for the same trajectory. Figures 2a(v-vii) are the CN, CO, and CO' internuclear distances. For the first 0.6 ps these distances oscillate about the equilibrium nitromethane values but then rapidly increase, indicating C-N bond dissociation. This is followed by an immediate decrease in R(C-O') to 1.421 Å [Fig. 2a(vii)] as $CH_3O'NO$ is formed. Perhaps the most dramatic indication of isomerization are the ONO, CNO, and CNO' angles as functions of time, shown in Figures 2a(viii-x). At the beginning of the trajectory, both CNO and CNO' angles oscillate about 117.1°. After isomerization, the CNO' angle oscillates about 32°, the equilibrium CNO angle in methyl nitrite [Fig. 2a(x)]. Finally, Figure 2a(xi) shows the CO'NO dihedral angle for the portion of the trajectory during which methyl nitrite exists. The newly formed methyl nitrite undergoes two complete rotations, and then settles into the cis conformation (0°) until dissociation occurs.

The mechanism of isomerization on PES 2 and PES 3 is concerted: as the C-N bond breaks, the C-O bond is formed. Because isomerization on both surfaces occurs in the same manner, we have chosen to illustrate this mechanism of isomerization and dissociation using PES 3. Figures 2b(i-iv) and 2b(v-xi) illustrate the weighting functions and geometries

of a trajectory which results in direct isomerization from nitromethane
to methyl nitrite and subsequent C-O bond scission to form CH_3 + NO_2.
Figures 2b(i-iv) show the weighting functions $S_{CH_3NO_2}$, $S_{CH_3ONO'}$,

$S_{CH_3O'NO}$, and $S_{CH_3+NO_2}$, respectively. The total energy of the system is

69.2 kcal/mol above the zero point energy of nitromethane. For the
first 4 ps of the trajectory [Fig. 2b(i)], the system is described
mainly by the nitromethane term. At 4 ps, $S_{CH_3NO_2}$ decreases to zero as

$S_{CH_3O'NO}$ rises to unity. The system is then described by the methyl

nitrite term for the next 6 ps, at which point the C-O bond rupture
occurs. Therefore, the weighting functions show the mechanism of
isomerization to be

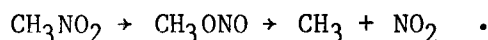$$CH_3NO_2 \rightarrow CH_3ONO \rightarrow CH_3 + NO_2 \quad .$$

Figure 2b(v-xi) show the same internal coordinates as in Fig. 2a
(v-xi) during the trajectory, confirming the analysis of the mechanism
given by the weighting function.

ENSEMBLE RESULTS: The curves in Figs. 3a and 3b are the
probabilities of Reactions III and II, respectively, as functions of
energy for the three potential energy surfaces. For the most part, the
probability of Reaction III increases with increasing energy for all
three surfaces. The behavior of the probability of Reaction II is quite
different among the three surfaces. The probability of Reaction II on
PES 3 decreases with increasing energy over the range of 64 to 100 kcal/
mol, but is less energy dependent at higher energies. The probabilities
of Reaction II on PES 2 and PES 1 are approximately constant over the
energy range studied.

The following ratios were calculated:

B_1 = Reaction II/(Reaction I + Reaction III);
B_2 = (Reaction II + Reaction III)/Reaction I; and
B_3 = Reaction II/Reaction III.

The ratio B_1 corresponds to the Wodtke, et al.,[2] measurements.
However, in determining the isomerization barrier, it seems that B_2
would be the ratio which needs to be considered. B_3 is the branching
ratio of the decomposition reactions of the methyl nitrite.

The branching ratios B_1 and B_2 are shown for PES 1 [Fig. 4a], PES
2, [Fig. 4b] and PES 3 [Fig. 4c]. The B_1 and B_2 for PES 1 have been
included for completeness; however, the statistical errors are very

large for the ratios due to the small number of isomerizing trajectories. The values of the branching ratio B_1 calculated for PES 2 show that the formation of $CH_3O + NO_2$ is essentially constant as a function of energy. However, B_2 increases with energy. The branching ratios calculated for PES 3 [Fig. 4c] exhibit very different behavior than those for PES 2, although the mechanism of isomerization is similar. Both the B_1 and B_2 curves initially decrease with increasing energy over the range 64 to 103 kcal/mol, indicating that isomerization is more probable at lower energies. At higher energies, both ratios are constants as functions of energy. Although the curves in Fig. 4c are significantly different than those in Fig. 4b, the probability of Reaction III on both PES 2 and PES 3 have the same energy dependence as shown in Fig. 3a. The difference in the curves in Fig. 4b and 4c is due to the difference in the behavior of the probability of Reaction II as a function of energy for PES 2 and PES 3 as shown in Fig. 3b. On PES 3, formation of $CH_3O + NO$ is strongly energy dependent at lower energies, whereas the same is not true for PES 2. The branching ratios which are calculated from the results for PES 3 most closely resemble the experimental values, and indicate that the barrier height of a tight transition state such as those described by PES 2 and PES 3 should be slightly larger than 47 kcal/mol.

The branching ratios B_3 are shown for PES 1 [Fig. 4d], PES 2 [Fig. 4e], and PES 3 [Fig. 4f]. Because so few trajectories resulted in isomerization and dissociation on PES 1, B_3 on this surface has large statistical error. However, the curves for B_3 for PES 2 and PES 3 (Figs. 4e and 4f, respectively) have similar energy dependences. At lower energies (below 100 kcal/mol), B_3 is larger than 1.0, indicating that upon isomerization, O-N bond scission in methyl nitrite is more probable than C-O bond scission. However, at higher energies, B_3 is approximately constant and less than unity, indicating that C-O bond scission is more probable following isomerization than O-N bond scission.

Two questions arise from the behavior of the reaction probabilities and the branching ratios:

a. Why does the C-O bond scission (Reaction III) occur at all energies and is more probable than O-N bond scission (Reaction II) in methyl nitrite at high energies, even though Reaction III requires 16 kcal/mol more energy?

b. Why is the probability of Reaction II essentially not energy dependent on PES 2, but decreases as a function of energy on PES 3?

Unimolecular decomposition of methyl nitrite under the experimental conditions of Wodtke, et al.,[2] would most likely result solely in O-N

bond breaking to form CH_3O + NO because of the large difference in the dissociation energies of the O-N and C-O bonds. However, the C-O bond that is formed when nitromethane isomerizes to methyl nitrite is very likely initially highly excited. That is, energy is selectively deposited in this bond in the isomerization reaction. The transfer of only a small amount of energy from other molecular modes will result in the breaking of the bond. Thus, this mode selectivity in the energy distribution of the nascent methyl nitrite leads to the rupture of the stronger C-O bond (preferentially at high energies) seen in our results.

The observed increasing probability of Reaction III with increasing energy is simply due to the increasing amount of energy that is available among the various modes for transfer to the excited C-O bond. Why then is Reaction II essentially constant as a function of energy for PES 2 but not for PES 3? This can be attributed to the difference in the barrier heights for isomerization from nitromethane to methyl nitrite for the two surfaces. The barrier height for isomerization on PES 2 is 55.1 kcal/mol and 47.6 kcal/mol for PES 3. If we assume that all of the energy of isomerization goes into the newly-formed C-O bond, then the C-O bond on PES 2 will have 7.5 kcal/mol more energy than the C-O bond on PES 3 because of the difference in the barrier heights. On PES 2 only 2 kcal/mol of energy must transfer from other modes to provide the energy needed to break the C-O bond (bond dissociation energy is 57 kcal/mol) and thus the reaction is essentially independent of the total energy. However, on PES 3 9.4 kcal/mol must flow into the C-O bond to break it. The higher energy requirement for C-O bond scission on PES 3 could explain the observed energy dependence of Reaction II on this surface.

Figure 5 shows the translational energy distribution of products from Reaction I and Reaction III for PES 3 at 69.2 [Fig. 5a], 115.3 [Fig. 5b], and 161 kcal/mol [Fig. 5c] above the zero-point energy of nitromethane. Note that the distributions of products for Reaction III are strikingly similar to those of Reaction I. Remarkable similarities are also apparent in the average translational and rotational energies of the products of Reactions I and III shown in Figs. 6, 7a, and 7b. Figure 6 is a plot of the average relative product translational energies for Reactions I, II, and III on PES 3 vs. energy. Figures 7a and 7b are plots of the average rotational energies of the methyl and nitrogen dioxide radicals, respectively, formed from Reactions I and III on PES 3. The similarities of the average rotational and translational energies of the products of these two reactions indicate that products from Reactions I and III might be very difficult to distinguish experimentally. The average translational energy of the products of Reaction II are, for the most part, larger than the average translational energy of the products of Reaction I and Reaction III at all energies studied. This is expected because the energy required for

Reaction II (O-N bond scission in methyl nitrite, 41 kcal/mol) is much less than that required for Reaction I or Reaction III (60 and 57 kcal/mol, respectively).

VI. SUMMARY: We have performed a classical dynamics study of the unimolecular decomposition of nitromethane. Three model potential energy surfaces were used in the study which differ mainly in the heights of the barrier to the isomerization of nitromethane to methyl nitrite. PES 1 has a very high barrier of 216.4 kcal/mol, PES 2 is 55.1 kcal/mol, the value suggested by Wodtke, Hintsa, and Lee,[2] and PES 3 has a barrier of 47.6 kcal/mol, which compares well with the predicted MINDO/3 value of Dewar and Ritchie.[12]

Two primary unimolecular decomposition pathways were suggested by Wodtke, et al.,[2] in interpreting experimental results:

$$CH_3NO_2 \rightarrow CH_3NO_2^{\dagger} \rightarrow CH_3 + NO_2 \tag{I}$$

$$CH_3NO_2^{\dagger} \rightarrow CH_3ONO^{\dagger} \rightarrow CH_3O + NO \quad . \tag{II}$$

An additional primary unimolecular decomposition pathway was observed in our trajectory results for all three potential energy surfaces,

$$CH_3NO_2^{\dagger} \rightarrow CH_3ONO^{\dagger} \rightarrow CH_3 + NO_2 \quad . \tag{III}$$

Reaction I occurs by the same mechanism for all three potential energy surfaces. The isomerization reactions, Reactions II and III, on PES 1 occur through a mechanism of dissociation of nitromethane to $CH_3 + NO_2$, followed by recombination of the radicals to form methyl nitrite. The probability of isomerization on this surface increases with increasing energy. The mechanism of isomerization for PES 2 and PES 3 is a concerted C-N bond breaking C-O bond forming process. PES 2, the surface which has the barrier height suggested by Wodtke et al., predicts a branching ratio at least an order of magnitude smaller than that determined from the experimental results. PES 3, predicts a branching ratio in reasonable agreement with that reported by Wodtke, et al.[2]

Although PES 2 and PES 3 differ by only a few kcal/mol in barrier heights and show the same mechanisms for isomerization and decomposition, the dynamics results are dramatically different. The energy-dependent branching ratios of isomerized reactions to C-N bond scission reactions for PES 2 increase with increasing energy. The same branching ratios for PES 3 decrease with increasing energies up to 103 kcal/mol, then remain constant at higher energies. There is a strong energy dependence of the probability of Reaction II for PES 3, whereas the probability of Reaction II is independent of energy on

PES 2. However, the probability of Reaction III increases with increasing energy for both PES 2 and PES 3.

Average translational energies of the products for Reaction II are higher than those of Reactions I and III at all energies. The product translational energy distributions, as well as the average translational and rotational energies of the products of Reactions I and III are similar for all three surfaces, indicating that the products of Reaction I and Reaction III might be indistinguishable in experiments.

Although the trajectory results suggest that the barrier height of the transition state should be slightly higher than 47 kcal/mol, several good quality ab initio calculations[4,5,6] predict a much higher barrier to isomerization for a tight transition state of the type described by the potential energy surfaces in this study. It is not known whether a much better electronic structure calculation would lower the energy value of this transition state by 20 to 30 kcal/mol. However, there have been suggestions that another transition state might exist which resembles a van der Waals complex.[7,8] Isomerization through this transition state would resemble the dissociation-recombination mechanism seen in the results for PES 1.

Enormous gains have been made previously in understanding the electronic structure and dynamics of this system. This study, however, brings up new questions and shows complexities regarding the primary decomposition pathways and mechanisms of dissociation. Therefore, a great deal remains to be done both in theory and with experiment before a definitive determination is made of how the unimolecular decomposition of nitromethane occurs.

REFERENCES

1. For a review of work on nitromethane, see J. Hershkowitz and
    B.M. Dobratz, US Army ARDEC Special Publication ARFSD-SP-89001,
    AD-E401-909, April, 1989.

2. A.M. Wodtke, E.J. Hintsa, and Y.T. Lee, J. Chem. Phys., Vol. 84,
    p. 1044, 1986: J. Phys. Chem., Vol. 90, p. 3549, 1986.

3. B.H. Rockney and E.R. Grant, J. Chem. Phys., Vol. 79, p. 708, 1983.

4. G.F. Adams, unpublished results.

5. M.L. McKee, J. Am. Chem. Soc., Vol. 108, p. 5784, 1986; Vol. 107,
    p. 1900, 1985.

6. C.F. Melius, private communication.

7. M.L. McKee, J. Phys. Chem., Vol. 93, p. 7365, 1989.

8. Roberta P. Saxon, private communication.

9. S.W. Benson, Thermochemical Kinetics, Wiley, New York, 1976.

10. W.D. Gwinn, R.J. Anderson and D. Stelman, Bull. Am. Phys. Soc.,
    Vol. 13, p. 831, 1968.

11. L. Batt and G.N. Robinson, The Chemistry of Amino, Nitroso, and
    Nitro Compounds and Their Derivatives, S. Patai, ed; Wiley, NY,
    1982.

12. M.J.S. Dewar and J.P. Ritchie, J. Org. Chem., Vol. 50, p. 1031,
    1985.

13. W.H. Miller and T.F. George, J. Chem. Phys., Vol. 56, p. 5668,
    1972.

14. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and
    E. Teller, J. Chem. Phys., Vol. 21, p. 1087, 1953.

15. L.M. Raff and D.L. Thompson, in Theory of Chemical Reaction
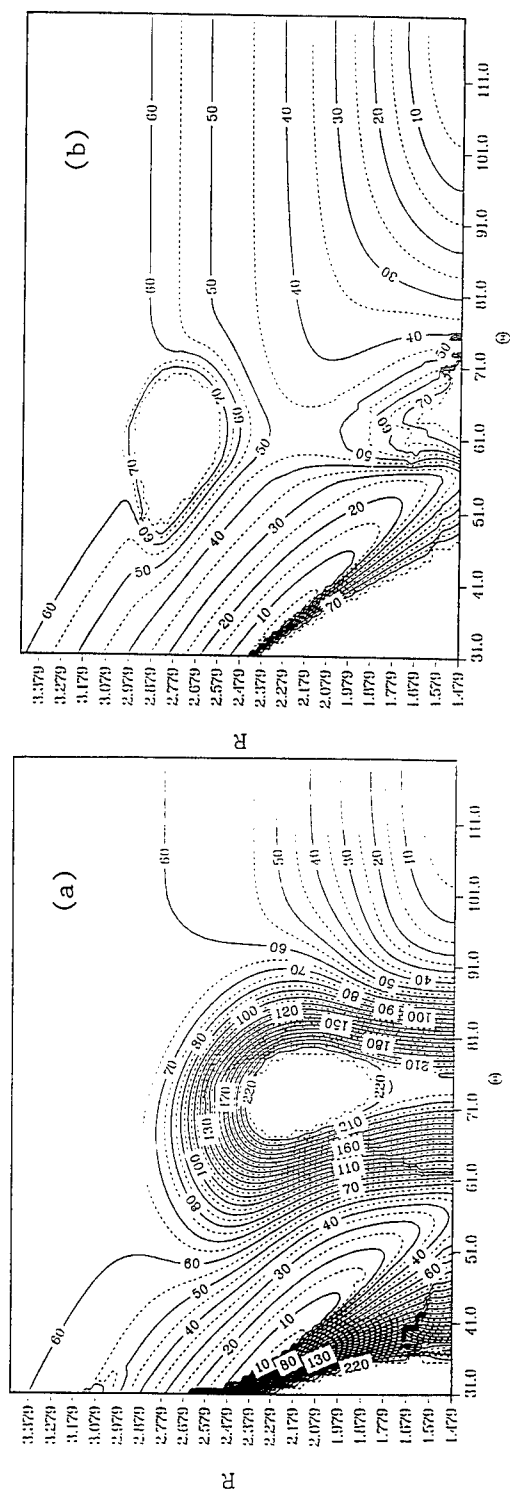   Dynamics, edited by M. Baer, Chemical Rubber, Boca Raton, FL, 1985.

Fig. 1. Contour plot of potential energy as a function of the C-N bond distance and CNO angle. Energy contour values are in kcal/mol, and shown at 5 kcal/mol increments. (a) PES 1 - contour values from 0.0 to 225 kcal/mol; (b) PES 3 - contour values from 0.0 to 75 kcal/mol.
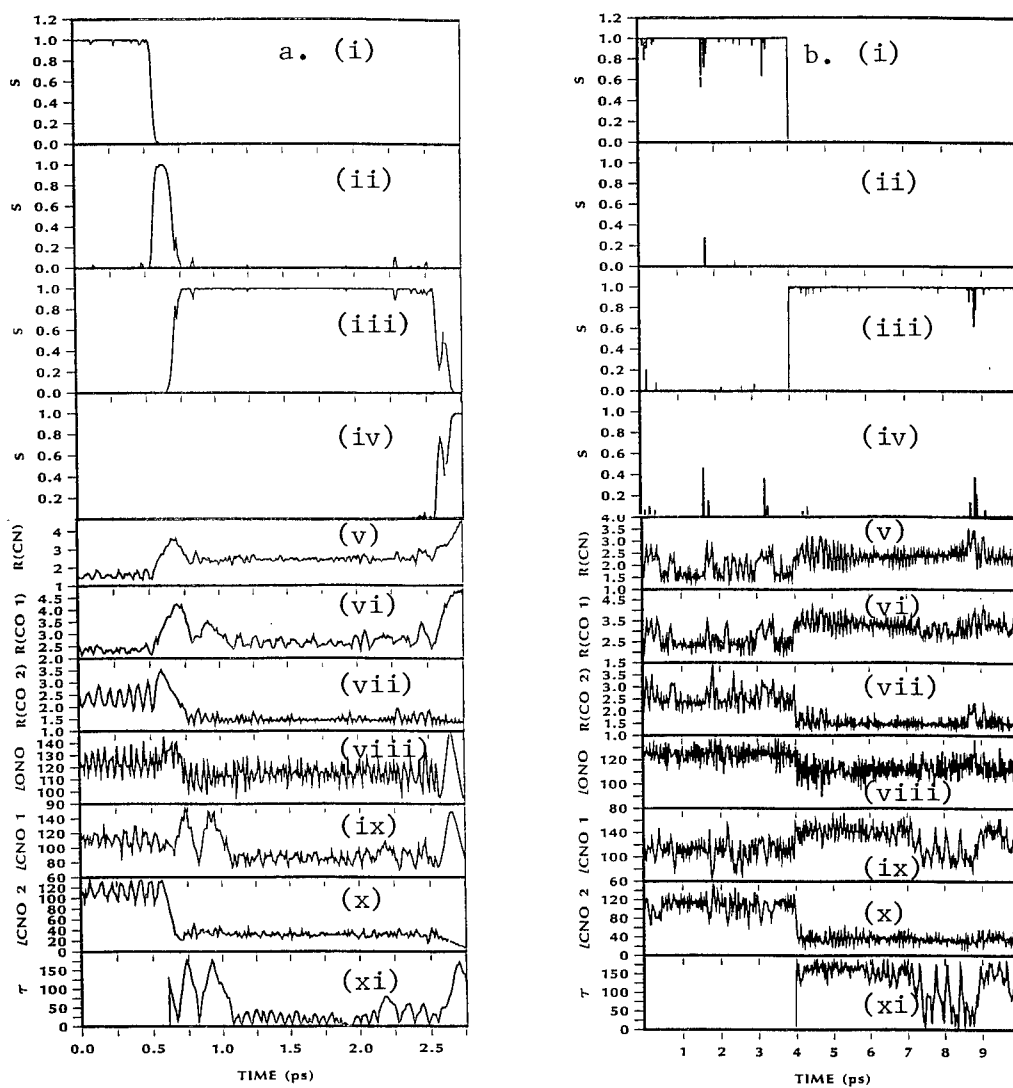
Fig. 2. A trajectory illustration for (a) Reaction II on PES 1, and (b) Reaction III on PES 3. See text, Section V.
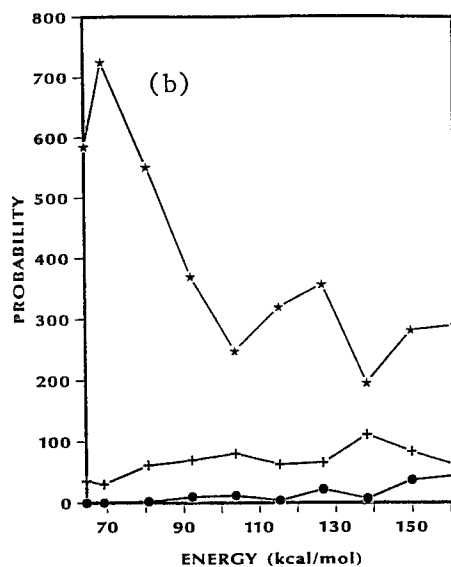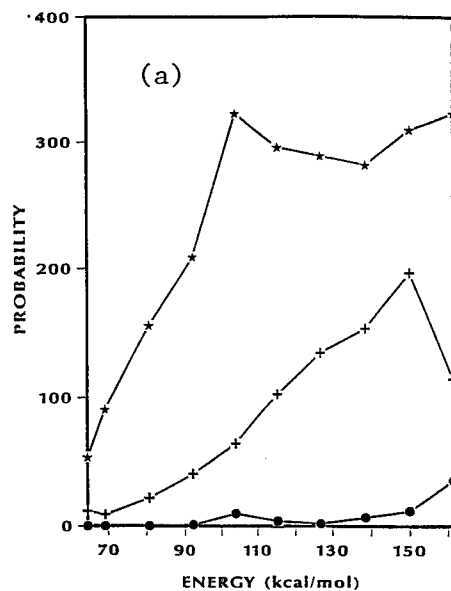
Fig. 3. Probability of Reactions (a) III and (b) II vs. energy for PES 1 (circles), PES 2 (crosses), and PES 3 (stars).
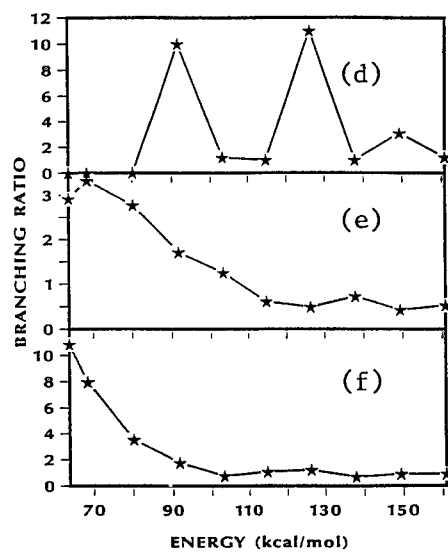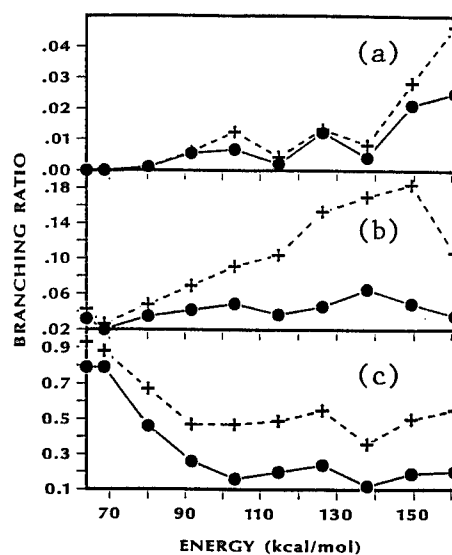
Fig. 4. Branching Ratios B_1 (circles) and B_2 (crosses) vs. energy for (a) PES 1, (b) PES 2, and (c) PES 3: and B_3 for (d) PES 1, (e) PES 2, and (f) PES 3.
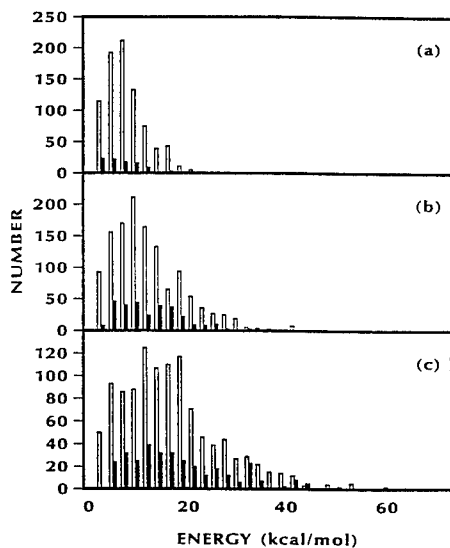
Fig. 5. Relative product distribution of translational energies for $CH_3 + NO_2$ from Rxns I (hollow bars) and III (filled bars) on PES 3 at energies (a) 69.2, (b) 115.3, and (c) 161 kcal/mol.
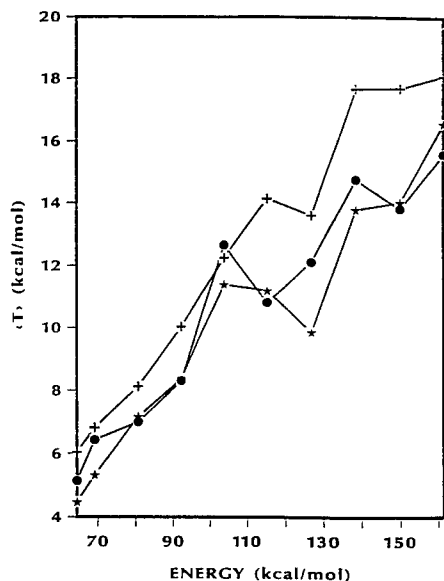

Fig. 6. Average relative product translational energies vs. total energy for Rxns I (circles), II (crosses), and III (stars) on PES 3.
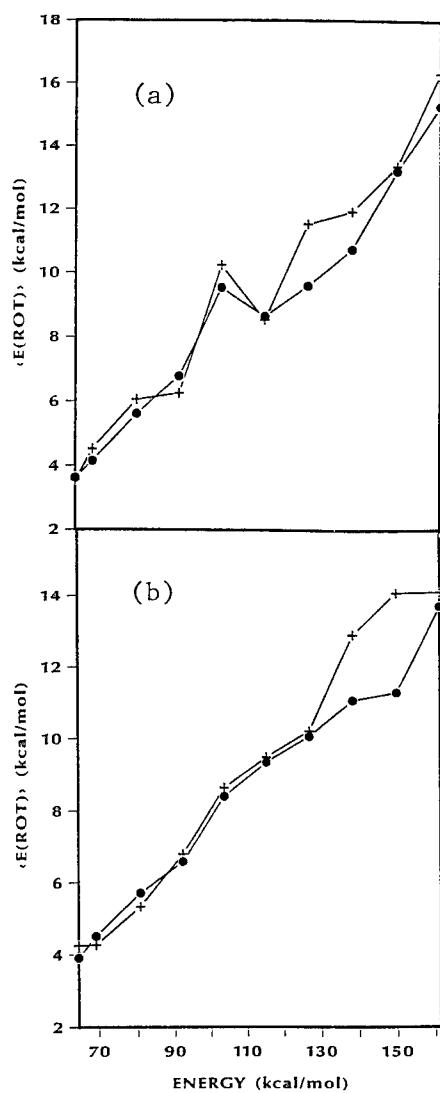
Fig. 7. Average relative product rotational energies vs. total energy for (a) CH_3, and (b) NO_2 from Reactions I (circles) and III (crosses) on PES 3.

GaAs Based Multiple Quantum Well
10 Micron Infrared Detectors (U)

*K. K. Choi, Dr., M. Dutta, Dr., P. G. Newman, Mr.,
W. D. Braddock, Mr.,
U. S. Army Electronics Technology and Devices Laboratory
Fort Monmouth, NJ 07703-5000

Recently, 10 micron infrared photoconductors have been successfully fabricated using GaAs/AlGaAs material system[1,2,3] instead of the more conventional HgCdTe material. The advantages of this technological innovation are substantial. The new devices are cheaper, faster, more reliable and flexible, and more easily integrated into supporting electronic circuits. They can be utilized in applications related to night vision, missile seeking, infrared optical communications and medical instrumentation. Arrays of these detectors may also be more easily fabricated than those based on the II-VI material system.

We have improved and extended the capability of these detectors in two important areas. First, we invented a 10 micron infrared transistor by incorporating a light sensitive structure in the emitter region of a hot - electron transistor, thereby improving its sensitivity in all temperatures. Second, we invented a voltage tunable multi-color detector using coupled quantum well structures. The multi-color detection capability of the detector greatly enhances the flexibility and reliability of the detector in optical communication and remote sensing.

The structure of an infrared phototransistor is shown in Figure 1. A typical device consists of a heavily doped ($n = 2 \times 10^{18} cm^{-3}$, 3000 Å) contact layer (1) as the emitter, a multiple quantum well structure (2) which comprises of 50 period of $Ga_{0.73}Al_{0.27}As$ undoped barrier (150 Å) and GaAs well ($n = 2 \times 10^{18} cm^{-3}$, 70 Å), a lightly doped ($n = 2 \times 10^{17} cm^{-3}$, 1500 Å) GaAs layer (3) as the base, an undoped $Ga_{0.73}Al_{0.27}As$ layer (700 Å) (4) as an electron energy filter, a heavily doped ($n = 2 \times 10^{18} cm^{-3}$, 6000 Å) GaAs layer (5) as the collector contact, and the entire device structure is grown on semi-insulating GaAs substrate. During the operation of the radiation detection, a bias is applied between layer (1) and layer (4) by a voltage source (7), the photocurrent generated is measured by a current meter (8).

In an experiment with a specified temperature, infrared radiation, from a glow-bar light source through a monochromator, is incident in a

45° angle relative to the superlattice layers via a polished face of the substrate. In this light coupling arrangement, the radiation with the electric vector component perpendicular to the superlattice layers will be absorbed (Figure 1) and generates photocurrent.
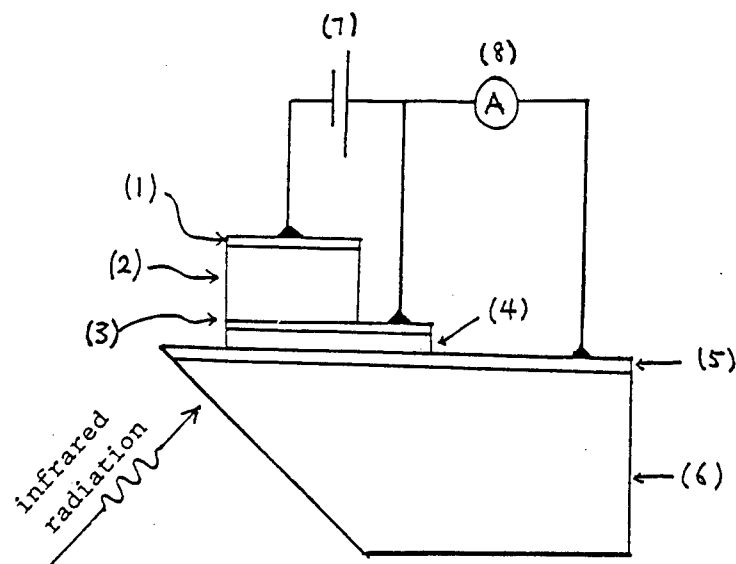

Figure 1. The cross section of the device and the electrical circuit.

The band diagram of the device is shown in Figure 2. When the device is under external bias V_E, the voltage between the emitter and the base, electrons with different energy will be injected into the base. The energy distribution of the hot - electrons in the base depends on the temperatures. The barrier between the base and the collector acts as an electron energy high pass filter. The electrons with energy higher than the barrier height of the filter overcome the barrier and are collected at the collector. On the other hand, the low energy electrons are collected at the base as base current. When the device is exposed to infrared radiation, electrons in the first subband E_1 of the multiple quantum well structure absorb optical energy $h\nu$ and are excited to the second subband. The photoexcited electrons then inject into the base as high energy electrons. These high energy electrons are then permitted through the filter and are collected as collector current. Therefore, by measuring the collector photocurrent, one can filter out a majority of the dark current which possesses lower energy. Since the electrical noise produced

by the dark current is the major source of noise in these devices, the detectivity of the detector can hence be increased compared with the conventional photoconductor without the energy filter.
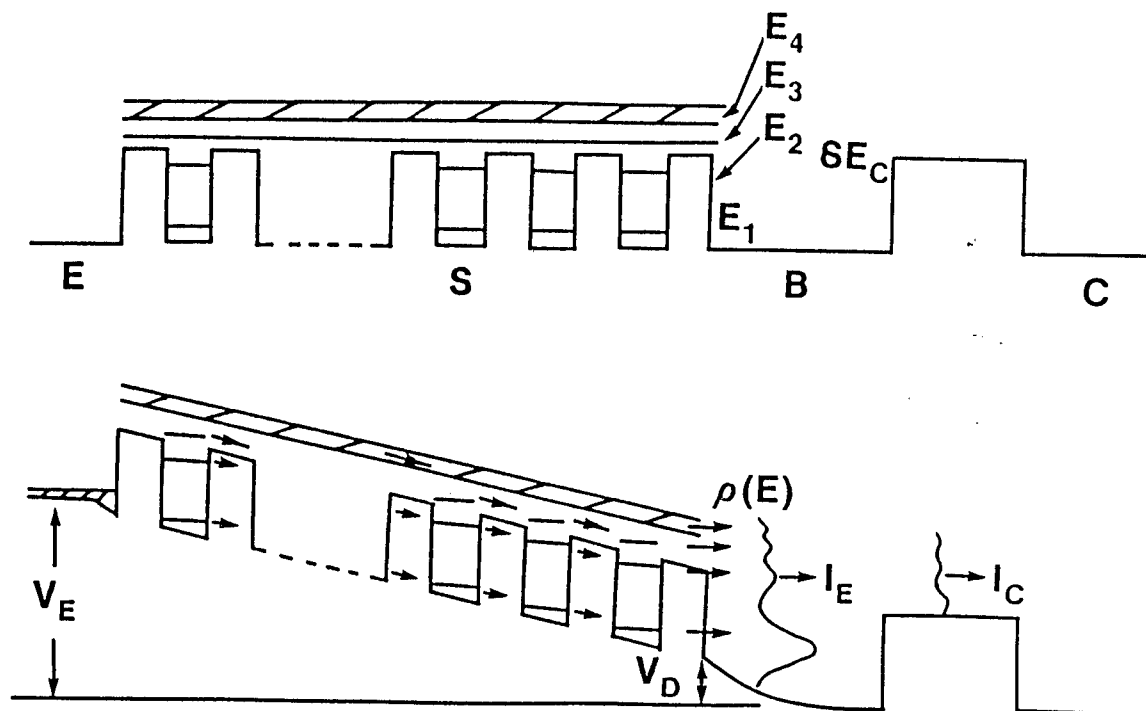
Figure 2. The band diagram of the device with and without bias.

In order to demonstrate that the photoelectrons are higher in energy than the dark electrons, we injected a monochromatic infrared radiation of 124 meV photon energy from a CO_2 laser into the device. Figure 3 shows the derivative (r) of the current transfer ratio (I_C/I_E) as a function of the external bias V_E at temperature of 10K with and without infrared radiation, where I_C and I_E are the collector current and the emitter current respectively. r can be shown to be directly proportional to the

313

electron energy distribution. Curve (0) is the measured r without light. The single peak observed at 4.2 V represents the dark electrons finally accepted into the collector at large enough external bias. Curve (1) and the curves in the insert are that with infrared light exposure. The curves demonstrate that the photoelectrons, being at higher energy, are accepted into the collector at lower external bias and form a photocurrent peak.
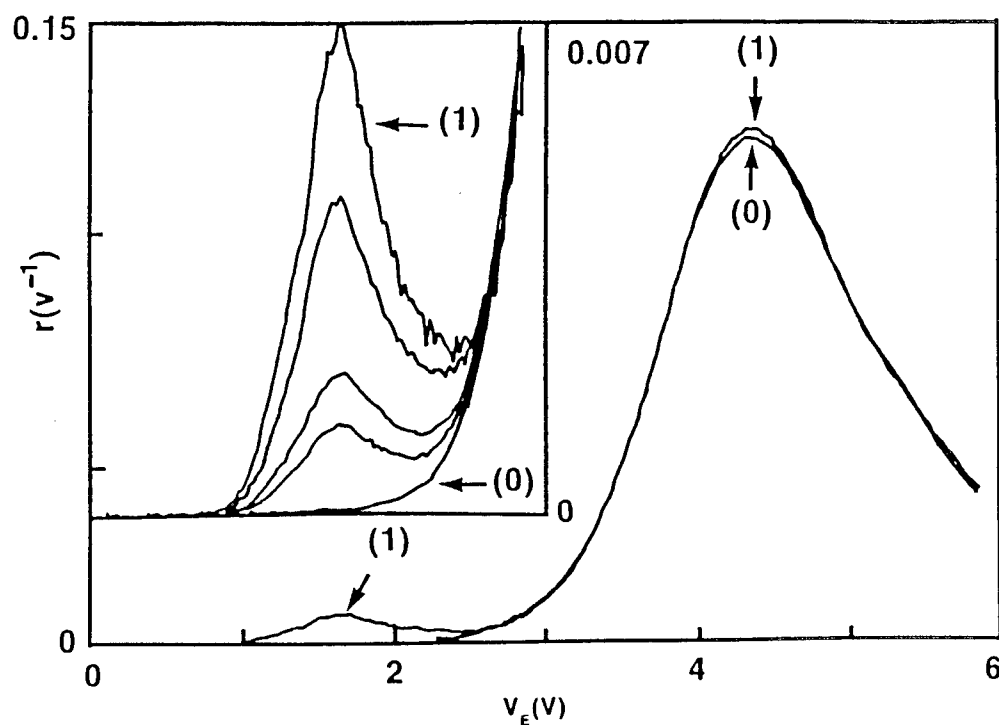

Figure 3. Curve (0) is the measured r without light. Curve (1) and the curves in the insert are that with different intensity of light illumination.

A demonstration of the utility of the filter is shown in Figure 4. Fig. 4(a) and (b) show the measured responsivity R of the device at two voltages, 3V and 3.7V respectively, with radiation of different wavelengths at temperature equal to 10 K using a glowbar-monochromator light source. It is evident that although at 3.7V, R is larger by a factor of 2.6, the noise level is much larger due to the larger portion of the dark current being accepted at the collector. By operating the device at lower voltage to filter out the dark current, the detectivity (D^*), which depends on the signal to noise level, actually increases by a factor of 2.8 when V_E is decreased from 3.7V to 3V.
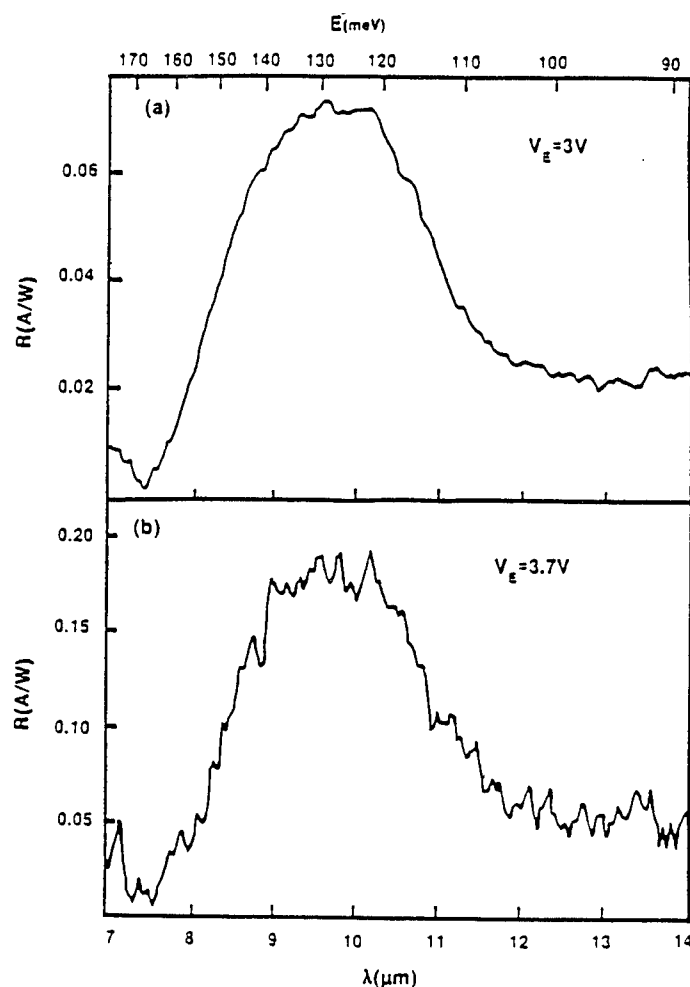

Figure 4. The measured responsivity at two different biases as a function of wavelength of the radiation.

The dark electron energy distributions at higher temperatures are shown in Figure 5, along with the energy distribution (the solid curve) under light exposure at 10 K. From this Figure, one notices that the peak of the thermal electrons appears at the higher voltage V_E than that for the peak of the photo-electrons when the temperature is below 100 K. Therefore, by operating the device at V_E just above the location of the photocurrent peak, the detectivity can be improved for all temperatures up to 100 K. At voltage equal to 2.4V, the device shows the detectivity, D^*, equal to 1.4×10^9 cm√Hz/W at 70 K, which is a factor of 3 better than the conventional GaAs detectors at this temperature. Even further improvement can be achieved if higher barriers are used to increase the operating voltage. The operating temperature can also be increased by using an electron energy band pass filter. These devices are currently under investigation.
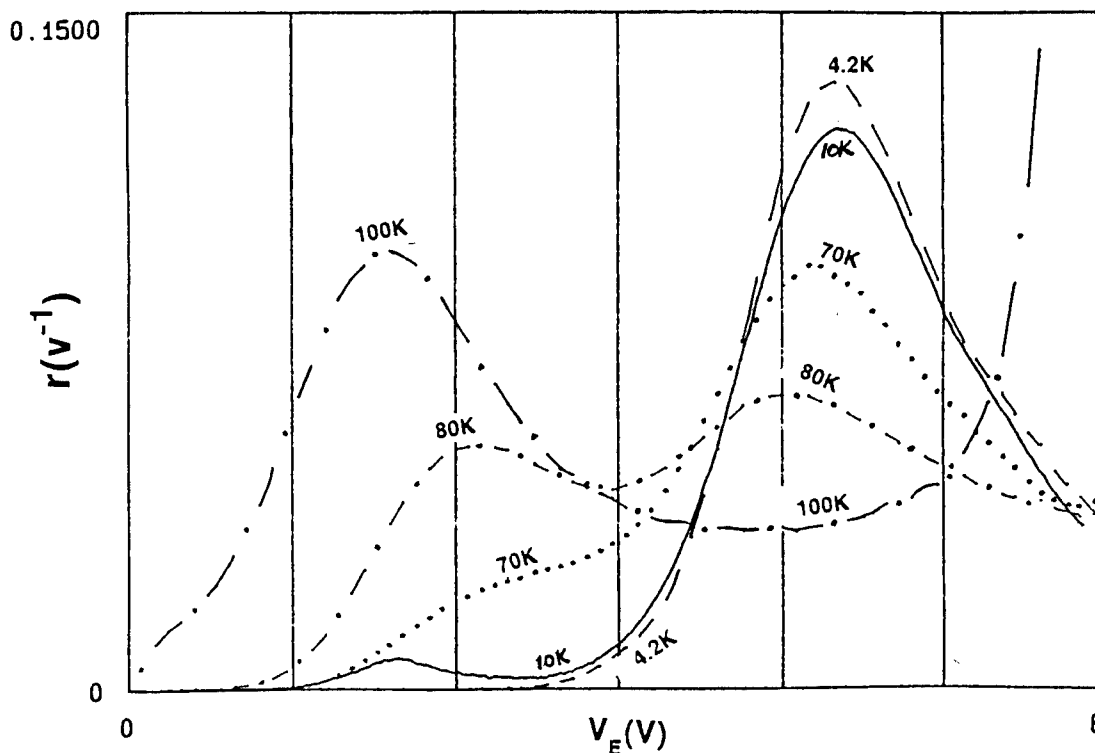

Figure 5. The measured r at different temperatures. The solid curve is with light, the other curves are without light.

For the multi-color photodetector, the device (Fig. 6(a)) consists of a semi-insulating substrate, and a semiconductor superlattice sandwiched between two contact layers. The device is biased by a voltage source through the two contact layers, and the photocurrent is sensed by either a current meter or by a series resistor.

The energy band structure of the device is shown in Fig.6(a). In a typical device structure, the thicker well (W_1), when it is isolated, contains two bound states (denoted E_1 and E_2 respectively in Fig.6(b)). On the other hand, the width of the thin well (W_2) is adjusted such that an isolated thin well will has an energy level (denoted E_1') very close to E_2. The level structure with different bias are depicted on the left hand side of Fig. 6(b), (c) and (d). When the two wells are separated by only a thin barrier (B_2), the electrons located in each well penetrate sufficiently into each other and change the level structure of each well. In the coupled quantum well structure, the new energy level structure, which is denoted E_1, E_2, and E_3 in the right hand side of Fig. 6(b), (c) and (d), exhibits different characteristics compared with the isolated wells. The most important difference is that in a simple single well unit, the electron wavefunction associated with each energy level has a definite parity, either even or odd. Since the intersubband optical absorption is an optical dipole transition, the electrons in the ground state, which are even in parity, can only be excited by the radiation to the odd parity states i.e. the second level, the fourth level ...etc. Hence, not all the levels can be used to detect radiation. On the other hand, for the present design of the coupled wells, each level is of mixed parity due to the breaking of the parity symmetry by the unequal well thickness. In this case, all the transitions from the ground state to all the excited states are allowed, leading to multi-color detection capability.

Another important feature of the coupled quantum well structure is the voltage tunability of the detector. For a simple single well design, the separations between the excited states and the ground state are extremely insensitive to the applied voltage. On the other hand, for the coupled quantum well structure, the separations between the levels are strongly affected by the applied voltage. It is because an applied voltage shifts the relative level positions originated from each well (Fig.6). It means that at each applied voltage, the frequencies of detection are different. Detectors of different frequencies can hence be obtained by simply changing the applied voltage.

Without infrared radiation illumination, when a small bias is applied, dark current is transported through ground state tunneling. When the voltage is increased, the electrons switch from ground state - ground state tunneling to ground state - excited state tunneling, thereby creating high field domains. The formation of high field domains can be observed in the oscillations in the dark current shown in Figure 7. When the device is exposed to infrared radiation, the electrons in the high field domain is excited from the ground state to one of the excited
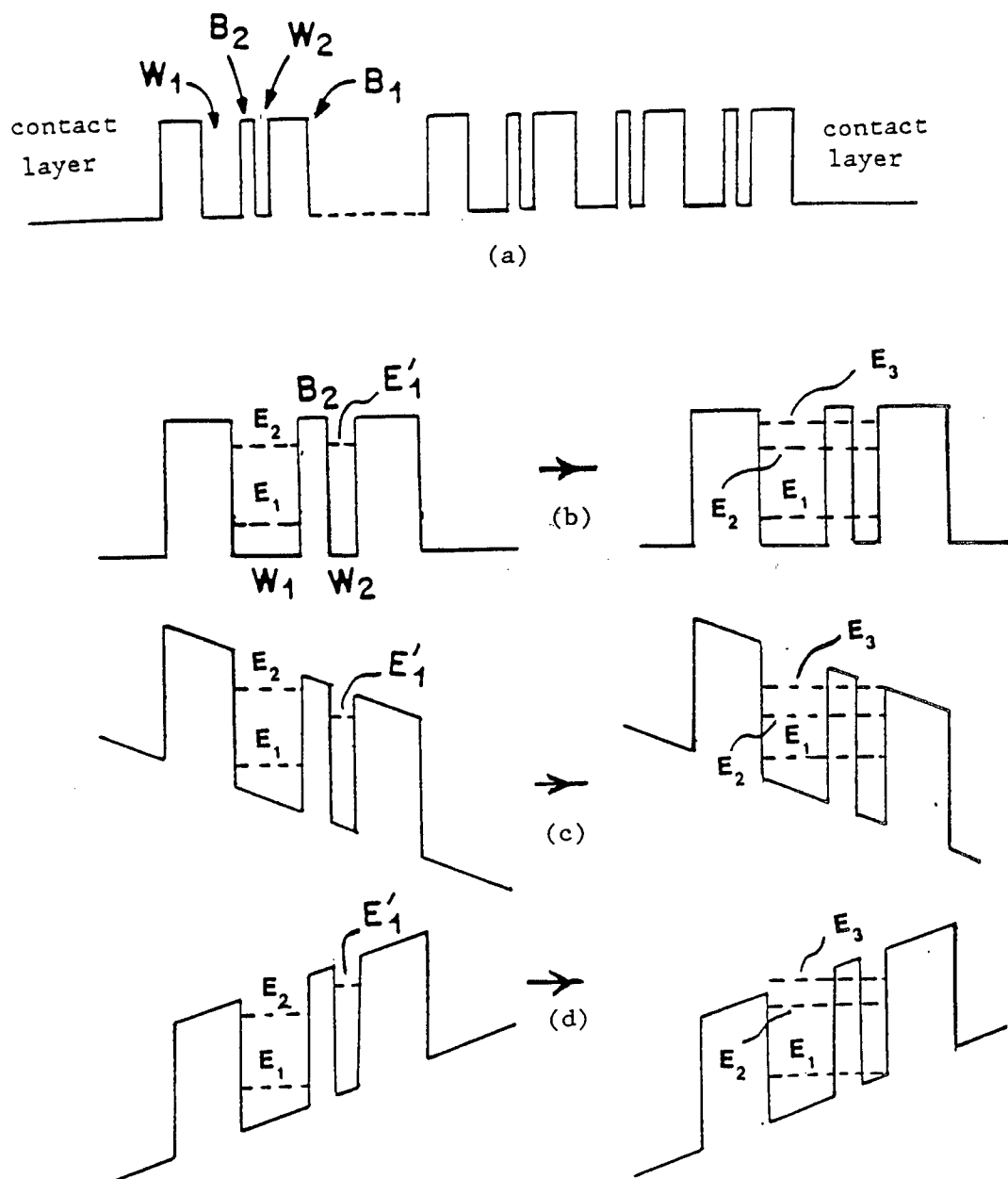
Figure 6. The band structure of the device (a). The band structures of one of the basic units under zero bias (b), forward bias (c) and reverse bias (d). The left hand side of the figure show the isolated level structures. The right hand side show the coupled level structures.

states. The electrons then tunnel out of the well forming hot electrons (inserts of Fig.8). Since the hot electrons can move more freely across the device than the tunneling electrons, the detector registers an increase in the current flow, by which the radiation is being detected.
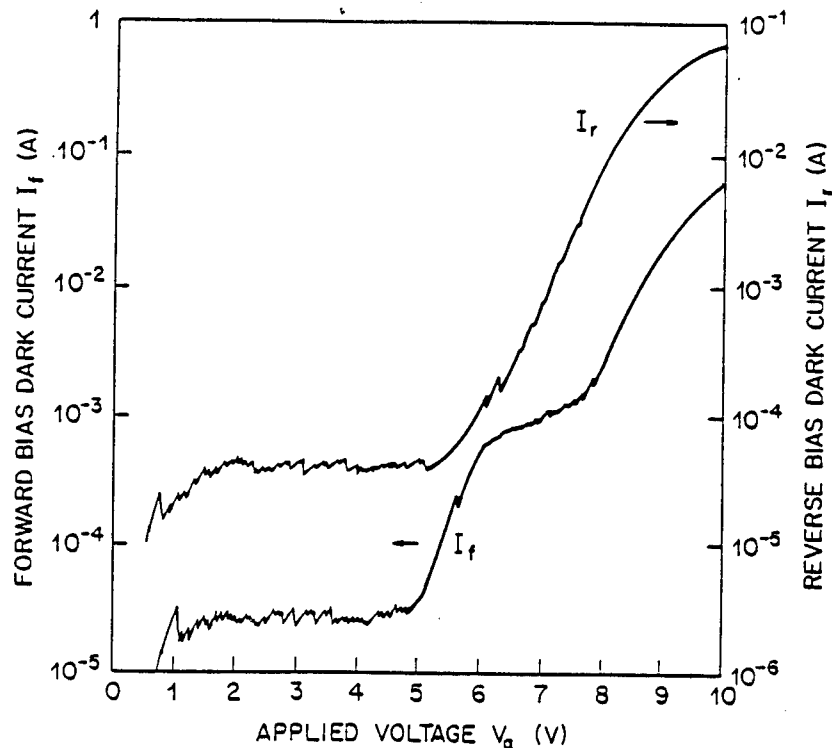

Figure 7. The measured dark current in both forward and reverse biases. The oscillations show the high field domain formation in the device.

The first sample is grown on semi-insulating GaAs substrate. It consists of 50 periods of 65-Å GaAs quantum well (doped with n = $1.0 \times 10^{18} cm^{-3}$), 40-Å undoped $Al_{0.25}Ga_{0.75}As$ barrier, 14-Å undoped GaAs well, and 150-Å undoped $Al_{0.25}Ga_{0.75}As$ barrier. The multiple quantum wells are sandwiched between the top (0.5 μm) and the bottom (1 μm) GaAs contact layers in which n = $1 \times 10^{18} cm^{-3}$. The resulting photoresponse as a function of radiation wavenumber ν is shown in Figure 8. In addition to the absorption curve shown, the solid curve shows the photocurrent when the detector is under reverse bias with the potential drop per period V_p = 63.5 mV, whereas the dashed curve shows the response under forward bias with V_p = 58.4 mV. Note that the positions of the photoresponse peaks depend on the applied voltage. The arrows shown in the figure indicate the expected peak positions calculated from the device structure.

Figure 8. The infrared absorption spectrum and the photocurrent for forward bias (dashed curve) and reverse bias (solid curve). The inserts indicate the device under different biasing conditions.

Figure 9 shows the photoresponse of another device design. The device consists of 50 periods of 72-Å GaAs (doped at $n = 1 \times 10^{18} cm^{-3}$), 39-Å $Al_{0.31}Ga_{0.69}As$, 20-Å undoped GaAs, and 154-Å undoped $Al_{0.31}Ga_{0.69}As$. The contact layers and the substrate are the same as the previous device. The data are taken at $V_p = 109mV$ for forward biasing (dashed curves) and $V_p = 123$ mV for reverse biasing (solid curve). The detectivity of the two detector is around 10^{10} cm/Hz/W at temperature of 10 K.
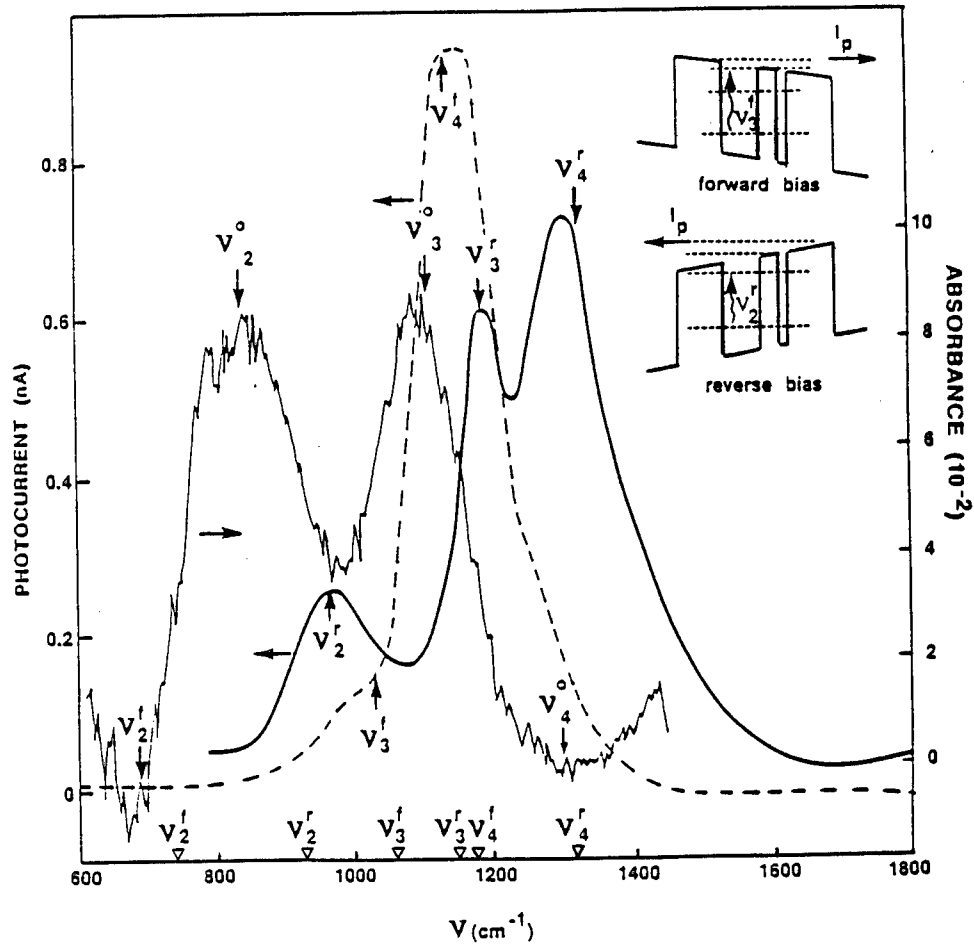
320
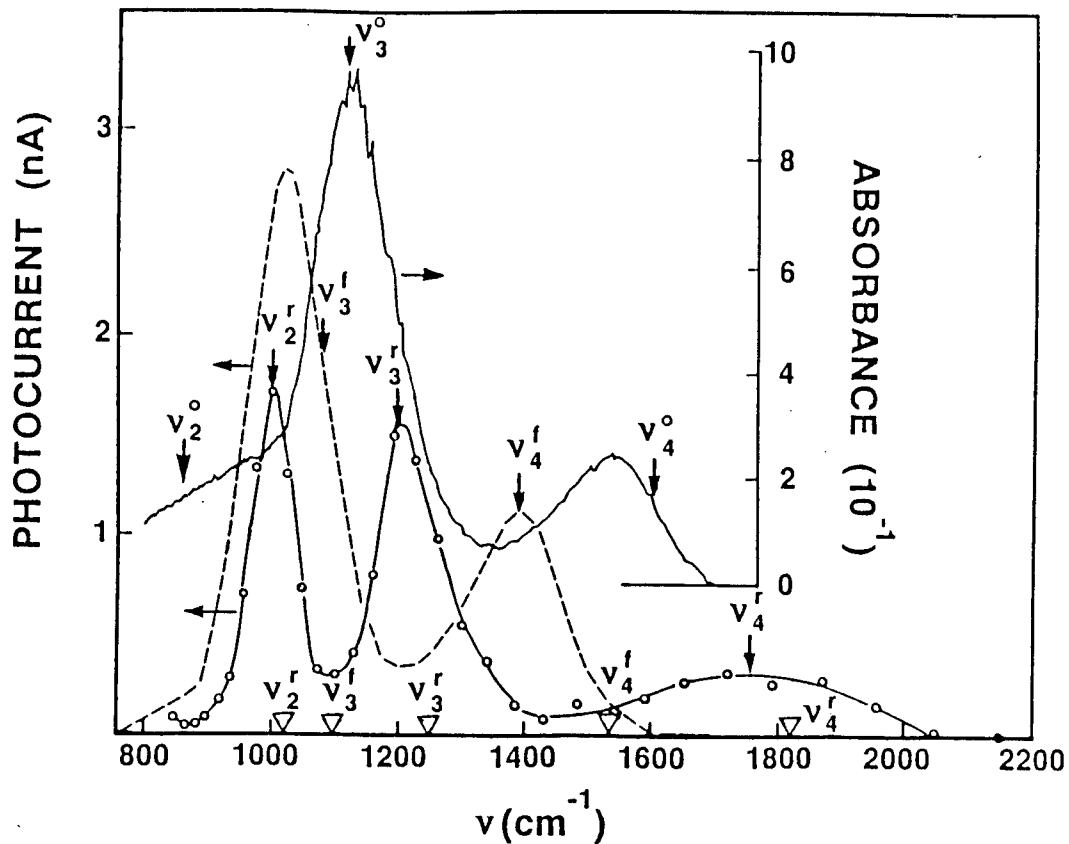
Figure 9. The infrared absorption spectrum and the photocurrent for forward bias (dashed curve) and reverse bias (solid curve).

In summary, a 10 micron infrared transistor based on GaAs material is invented. The transistor contains a quantum barrier electron energy filter in its collector. The energy filter reduces the electrical noise produced by the dark current, and hence increases the detectivity of the transistor. The detectivity of the present transistor is 1.4×10^9 cm/Hz/W at 70 K, three times the sensitivity of conventional GaAs photoconductors at the same temperature.

A multi-color infrared photoconductor is invented for optical communication and remote sensing. With its unique features of multi-color sensing and wavelength tunability, the detector is more flexible and reliable than the conventional GaAs device.

References
1. B. F. Levine, K. K. Choi, C. G. Bethea, J. Walker, and R. J. Malik, Appl. Phys. Lett. $\underline{50}$, 1092 (1987).
2. K. K. Choi, B. F. Levine, C. G. Bethea, J. Walker, and R. J. Malik, Appl. Phys. Lett. $\underline{50}$, 1814 (1987).
3. B. F. Levine, G. Hasnain, C. G. Bethea, and Narsh Chand, Appl. Phys. Lett. $\underline{54}$, 2704 (1989).

Advances in Modeling Combustion Processes
in Liquid Propellant Guns (U)

*Terence P. Coffee, Dr., Gloria P. Wren, Mrs.
Walter F. Morrison, Dr., and Paul G. Baer, Mr.
U.S. Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland

INTRODUCTION

In a regenerative liquid propellant gun (RLPG) a liquid monopropellant is injected at high velocity through an annulus into a combustion chamber where it ignites and burns. The liquid jet encounters a hot, dense, turbulent, high pressure environment of reacted and partially reacted products. Subsequently, the injected liquid undergoes an intricate series of processes including atomization, heating, vaporization, diffusion and turbulent mixing, and chemical reactions.

Although interior ballistic models of the RLPG over the past five years have been shown to be capable of accurately describing the mechanical response of the system, such as piston and projectile motions, descriptions of the breakup and combustion of the liquid jet in the combustion chamber have proven less satisfactory, involving empirical parameters which are dependent on gun caliber and configuration. In general, it has been assumed for simplicity that the process can be described in terms of the liquid propellant breaking into droplets whose diameters can be represented as functions of piston position or chamber pressure. Subsequent combustion proceeds according to a pressure-dependent burn rate law. The primary motivator for this description of the liquid propellant combustion process is the simulation of the time delay in energy release of the liquid propellant, referred to as accumulation, in the startup regime of the ballistic cycle.

The empirical description of jet breakup has been shown to provide accurate gun simulations and, in fact, to be reasonably predictive in terms of maximum pressure and muzzle velocity, within guns of the same caliber. However, it has not been possible until now to identify a unifying set of parameters which describe breakup and combustion among calibers, specifically among experimental data from 30mm, 105mm, and 155mm gun firings. The data from all three of these fixtures have been extensively compared to computer simulations to better understand the RLPG process.[1-3] This paper details significant progress toward

323

developing a jet breakup and combustion model which compares well with experimental data from all three calibers.

The design of interest, a Concept VIC RLPG, is shown in Figure 1. An external solid or liquid propellant igniter venting into the combustion chamber initiates the ballistic cycle. The chamber pressure forces both the control and injection pistons rearward. Liquid propellant is injected from the liquid reservoir through the annulus between the pistons into the combustion chamber where it burns, accelerating the projectile. Typically, the startup regime of the RLPG process is characterized by accumulation of liquid propellant in the combustion chamber in an unreacted or partially reacted state. Combustion models have been implemented to simulate this delay in energy release.
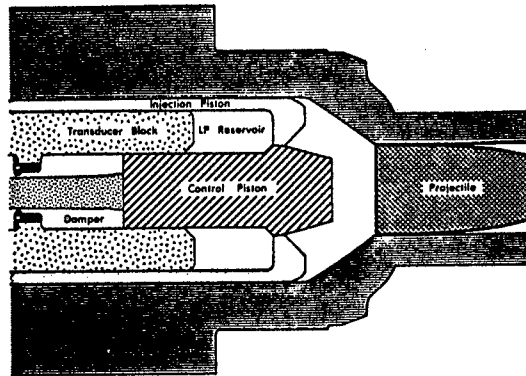

Figure 1. A Concept VIC Regenerative Liquid Propellant Gun

The most unsatisfactory aspect of the interior ballistic simulations has been the combustion model. The mean diameter of the droplets in the combustion chamber is generally input as a function of pressure. The droplets then combust according to a pressure-dependent burn rate law based on strand burner measurements.[4-5] An optimum droplet profile to match the chamber pressure-time history must be derived for each caliber and charge size. The droplet profiles for different initial conditions, and especially for different calibers, vary substantially in magnitude, although their qualitative behavior is similar.

In this paper we discuss models which appear to accurately predict the droplet profiles for all cases studied. A total of eleven different experimental cases in the 30mm, 105mm and 155mm gun firings have been considered to represent the range of all current Concept VIC experimental data. However, this report will focus on the four repeatability series within the three calibers. These were sets of ten shots in which the initial conditions were carefully controlled to be as identical as possible. In general, the guns demonstrated repeatability equivalent to solid propellant gun performance. Although not shown, the jet breakup and combustion models described in this report provide good simulations for all eleven cases.

PREVIOUS METHODS

The lumped parameter model consisting of ordinary differential equations has been described previously.[6-8] One option built into the interior ballistic code is an assumption of instantaneous propellant combustion. The assumption is made that the liquid propellant releases all its energy immediately upon injection into the combustion chamber. This is the limiting case of very rapid combustion. Table 1 shows the muzzle velocities predicted using this assumption with results given for the four repeatability series.

Table 1. Experimental Muzzle Velocities for Repeatability Series Compared to Simulation				
Gun	30mm	105mm	155mm	155mm
Charge (liters)	.23	2.0	2.2	5.2
Exp Muzzle Vel. (m/s)	776	666	393	586
Assuming Instantaneous Combustion				
Model Muzzle Vel. (m/s)	757	654	407	595
Percent Difference	-2.5	-1.8	3.6	1.5
Using the 30mm Droplet Profile				
Model Muzzle Vel. (m/s)	776	663	389	596
Percent Difference	0.0	-0.5	-1.0	1.7
Using the New Droplet Profile				
Model Muzzle Vel. (m/s)	767	658	399	593
Percent Difference	-1.2	-1.2	1.5	1.2
Using the Sensitive Time Lag Model				
Model Muzzle Vel. (m/s)	764	655	399	592
Percent Difference	-1.5	-1.7	1.5	1.0

Figures 2-5 show the experimental chamber pressures (solid line) for the four repeatability series, compared to a simulation using instantaneous combustion (dotted line). Although the predicted maximum pressure is within 8 percent of the experimental value in the worst case (30mm), the simulations using the instantaneous combustion model show poor agreement with the shape of the experimental pressure profiles. The simulations show the primary pressure rise occurring early due to the unrealistically high propellant gas generation rate at low pressure, indicating that liquid accumulation is important in these shots.
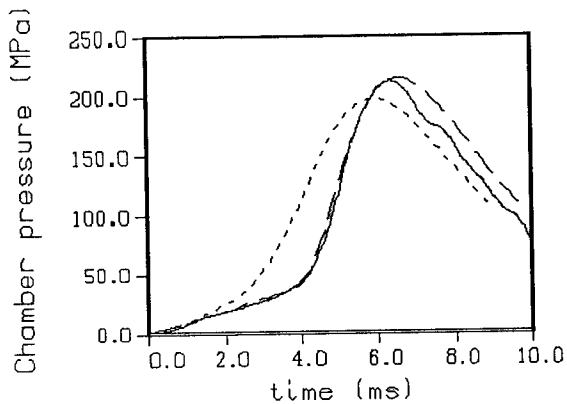
Figure 2. Combustion Chamber Pressure. 30mm Gun. Experiment (line). Instantaneous Combustion Model (dot). 30mm Droplet Profile (dash).
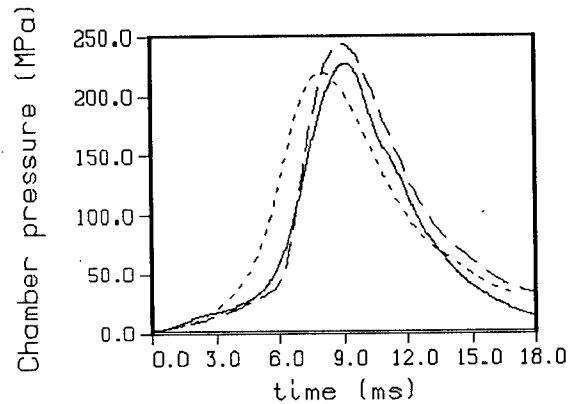
Figure 3. Combustion Chamber Pressure. 105mm Gun. Experiment (line). Instantaneous Combustion Model (dot). 30mm Droplet Profile (dash).
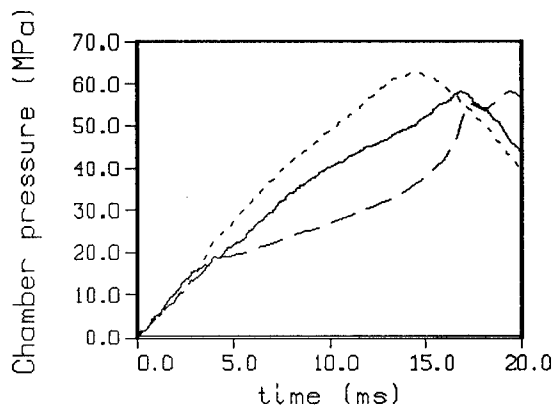
Figure 4. Combustion Chamber Pressure. 155mm Gun - 2 Liter Charge. Experiment (line). Instantaneous Combustion Model (dot). 30mm Droplet Profile (dash).
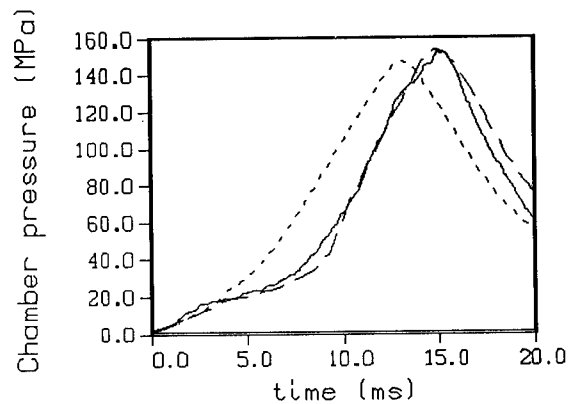
Figure 5. Combustion Chamber Pressure. 155mm Gun - 5 Liter Charge. Experiment (line). Instantaneous Combustion Model (dot). 30mm Droplet Profile (dash).

One method of including the effect of accumulation is to derive a droplet diameter profile as a function of pressure for each case.[1-3] In general, the use of a droplet diameter profile results in accurate simulation of the experimental chamber pressure profiles. However, the droplet diameter profiles for the four cases in Table 1 vary considerably, as shown in Figure 6. Although the droplet diameters decrease with pressure in all cases, the authors were unable to develop a formulation which generates the desired droplet diameters in all four cases.

326

Figure 6. Empirical Droplet Diameter Profiles. 30mm (line). 105mm (dot). 155mm - 2 Liter Charge (dash). 155mm - 5 Liter Charge (dot-dash).
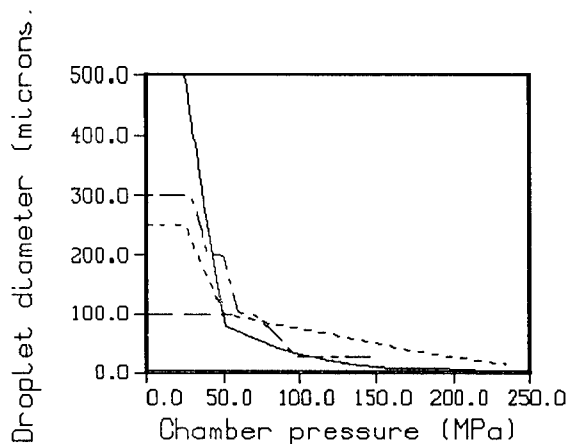
As an alternative to the instantaneous combustion model and the individual droplet diameter profiles, a droplet profile derived for one case may be applied to the simulation of other cases. The introduction of a droplet profile from another case does improve the overall interior ballistic simulation in comparison to the instantaneous combustion model. As an illustration, the droplet profile derived for the 30mm case is applied to the other three cases. The simulated muzzle velocities are presented in Table 1. Agreement is, of course, excellent for the 30mm case, but agreement is also improved for the other three cases. Muzzle velocity predictions are within the two percent range.

Figures 2-5 show the comparison of pressure versus time curves for these simulations (dashed lines) with experimental data. Agreement in Figure 2 is quite good, as expected. Disagreement after peak pressure is in part due to thermal drift in the pressure gauge. In the 105mm case (Figure 3), the position of the rapid pressure rise is approximately correct, but the slope of the curve is too large. Nevertheless, this is a credible simulation. In the 155mm, 2 liter case (Figure 4), the simulation is quite poor. Since the shot is a very low pressure firing, only the initial portion of the 30mm droplet diameter profile is used. For the higher pressure 155mm case (Figure 5), overall agreement is improved, although the initial part of the pressure rise (5 to 10 ms) is not simulated well. The maximum difference in the peak pressures between experiment and simulation is about six percent (105mm case).

MODELING ASSUMPTIONS

Our goal is to obtain a more fundamental model of the injection and combustion processes. The liquid propellant is injected as an annular jet at high velocity into a hot, high density gas. Both the velocity and the thickness of the jet change rapidly with time. The jet is assumed to break up into droplets. These drops can break up further or they can coalesce. The droplets are heated and eventually ignite. The droplets then combust and release the propellant energy. A number of assumptions

are required to make the problem tractable.

We assume that the liquid propellant instantaneously forms droplets as it enters the combustion chamber. Under gun conditions (very high velocities), the jet is expected to be in the atomization regime, i.e., droplets begin forming immediately upon entry into the chamber. While some liquid will remain in the jet core, the amount is assumed to be small (see below).

The droplets are assumed to begin to combust instantaneously, i.e., droplet heating and ignition is ignored. The droplets are assumed to burn according to an experimental, pressure dependent, linear regression rate.[4-5] The gas generation rate then depends only on the pressure and the total surface area of the droplets. The ignition systems in the VIC configurations produce a highly turbulent, hot, high pressure environment in the gun chamber; therefore, this should be a good approximation.

The liquid propellant burning rate has only been measured up to about 100 MPa. There is some evidence that there is a break in the slope of the burning rate around 95 MPa,[5] however, so the regression rate is not well known at high pressures.

Essentially all studies of jet breakup have been carried out under steady state conditions. We assume that steady state models still apply under gun conditions. Injection velocities and chamber conditions change rapidly during a gun firing. However, if the jet breakup is rapid enough, this will not be an important effect.

When a jet breaks up, a distribution of droplet sizes will be created. We assume that all the droplets created are of the same size, representing a mean diameter. The droplet diameter is chosen to be the Sauter mean diameter, which preserves both volume and surface area of a droplet distribution.

Jet breakup depends heavily on the relative velocity between the liquid and the gas. However, in the simulation the chamber gas is assumed to be stagnant, since the actual gas flow pattern in the chamber is unknown.

Once the droplets are formed, they are assumed not to interact with one another. This is probably the weakest assumption. The liquid propellant spray in the gun is very dense. Droplet combustion is known to depend upon neighborhood conditions.[9] Also, droplets could coalesce into larger droplets. Since the injection velocity is increasing over the first part of the firing cycle, later injected liquid could easily catch up to earlier liquid and form a dense liquid "clump". This effect is ignored only because we have no way of taking it into account.

INJECTION MODELS

Weiss and Worsham[10] performed a series of experiments in which molten wax was injected through circular holes into steady high velocity airstreams. The experimental droplet diameter could be correlated quite well by the relation

$$d = .5978 \ V^{-4/3} \ v^{1/12} \ D^{1/6} \ \sigma_L^{5/12} \ \mu_L^{1/3} \ \mu_G^{1/12}$$

$$\rho_L^{1/6} \ (1/\rho_G + 1/\rho_{G0}) \ , \tag{1}$$

where V is the relative velocity between the gas and liquid (cm/s), v is the absolute velocity of the liquid (cm/s), D is the injection hole diameter (cm), σ_L is the surface tension of the liquid (dynes/cm), μ_L is the liquid dynamic viscosity (poise), μ_G is the gas dynamic viscosity (poise), ρ_L is the liquid density (g/cc), ρ_G is the gas density (g/cc), and ρ_{G0} is the gas density at atmospheric pressure and room temperature (g/cc). The diameter reported is the mass mean diameter of the measured droplets. In these experiments the liquid velocity was much smaller than the gas velocity, so the contribution of the absolute liquid velocity term is negligible. Weiss and Worsham concluded that the atomization occurs by direct action of the airstream on the exposed liquid surface. Therefore, the relative velocity V between liquid and gas is of primary importance.

A later report by Wolfe and Andersen[11] presents a theoretical derivation of liquid breakup. While presented in terms of droplet breakup, the same general ideas apply to jet breakup. As in the description above, aerodynamic forces strip off blocks of liquid. However, Wolfe and Andersen also take into account the curvature of the droplet. They obtain the formula

$$d = 5.1426 \ V^{-4/3} \ D^{1/6} \ \sigma_L^{1/2} \ \mu_L^{1/3} \ \rho_L^{-1/6} \ \rho_G^{-2/3} \ , \tag{2}$$

where D is the diameter of the original large drop. Since this is a curvature effect, the outer diameter of the annular jet should be used, rather than the hydraulic diameter of the jet.

The formulas are similar. We would like to know if they are likely to be valid under gun conditions. The experimental velocities, viscosities, and surface tension studied are similar to those expected in the gun. The diameter of the gun annular jet is substantially larger than the circular jets used for the above correlations. The gas density in the gun starts small, but becomes 1-2 orders of magnitude larger than the cases studied by the above authors. Due to the difference in conditions, it was felt better to extrapolate a theoretical formulation rather than the purely experimental correlation of Weiss and Worsham. The Wolfe and Andersen formula is chosen since it takes into account the diameter of the jet, which varies substantially as the gun caliber is changed.

RESULTS - INJECTION MODEL

The Wolfe and Andersen formula was implemented in the gun code. The code predicts small droplets, and the resulting pressure-time curve is

very close to that resulting from the instantaneous combustion model. Although the simulation is disappointing, the standard steady state breakup models may not apply under the highly transient, very dense conditions of the gun. Alternate assumptions will be examined in a following section on intact jet core models.

However, it is still possible that the functional relationship is a reasonable approximation to gun conditions. Thus, a scaling factor is introduced and benchmarked against the 30mm case. If the droplet diameter is multiplied by 30, a reasonable approximation to the chamber pressure is obtained (Figure 7, dotted line). The agreement is not impressive, but the simulation does reflect the correct shape and magnitude of the experimental curve.

The model is then applied to the other three cases with the same scaling factor of 30. The agreement in all three cases is very good with no additional fine tuning of the injection model (Figures 8-10, dotted line). The agreement in muzzle velocity is also quite good (Table 1). The largest disagreement in the maximum chamber pressure is only 2.4 percent (30mm case).

Figure 7. Combustion Chamber Pressure. 30mm Gun. Experiment (line). Droplet Injection Model (dot). Sensitive Time Lag Model (dash).

Figure 8. Combustion Chamber Pressure. 105mm Gun. Experiment (line). Droplet Injection Model (dot). Sensitive Time Lag Model (dash).

The results are encouraging. All four cases are well represented by the same droplet model with the same value for the single adjustable parameter.

In general, the injected droplet diameter is nearly the same as the average droplet diameter in the chamber (Figure 11). During the early injection process, the injected droplet diameter decreases more rapidly than the diameter of the droplets in the chamber, which is decreasing due to combustion. As the injected droplet diameter levels off, this relationship is reversed.
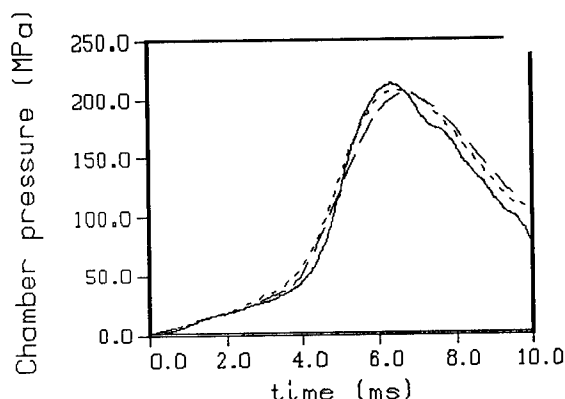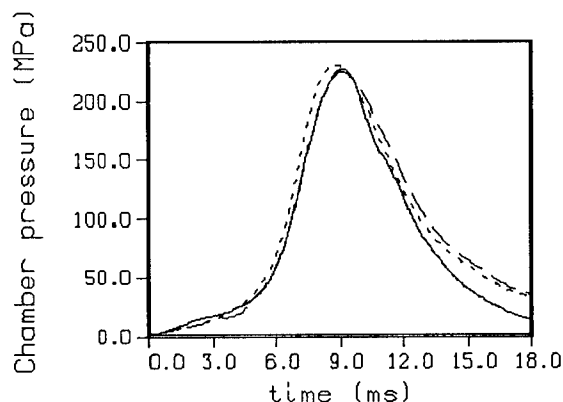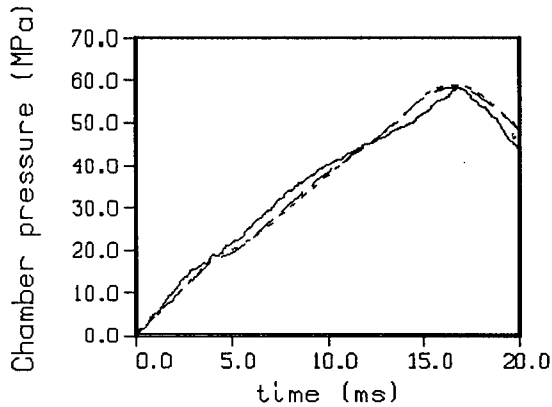
Figure 9. Combustion Chamber Pressure. 155mm
Gun - 2 Liter Charge. Experiment (line).
Droplet Injection Model (dot). Sensitive
Time Lag Model (dash).
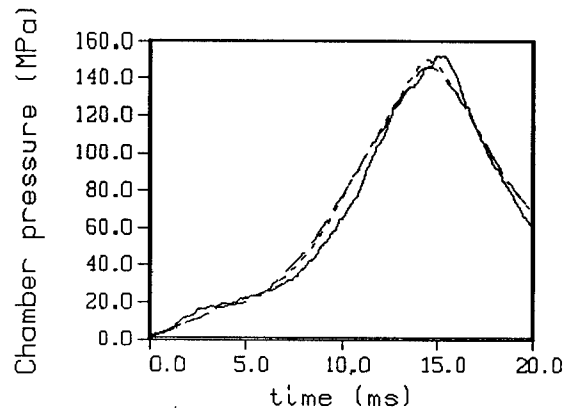
Figure 10. Combustion Chamber Pressure. 155mm
Gun - 5 Liter Charge. Experiment (line).
Droplet Injection Model (dot). Sensitive
Time Lag Model (dash).

In Figure 12, the new injected droplet diameter profiles for the four cases in Table 1 versus the chamber pressure are compared. The diagram is the projection of the function of multiple variables, Eq. (2), into the chamber pressure-droplet diameter plane. In all four cases, as pressure increases, the droplet diameter decreases until maximum chamber pressure is reached. After maximum pressure, liquid injection continues. As chamber pressure decreases from its maximum in each case, droplet diameter increases slightly.
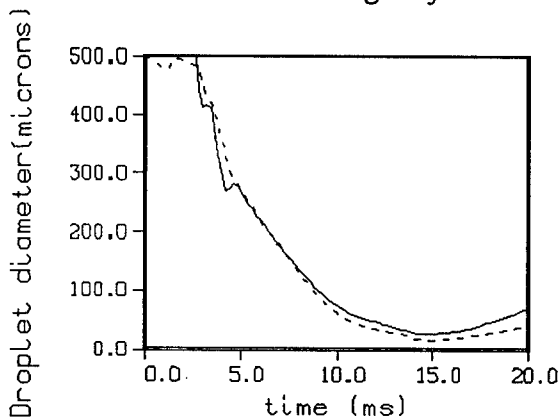
Figure 11. Droplet Injection Model. 155mm
Gun - 5 Liter Charge. Injected Droplet
Diameter (line). Combustion Chamber
Droplet Diameter (dot).
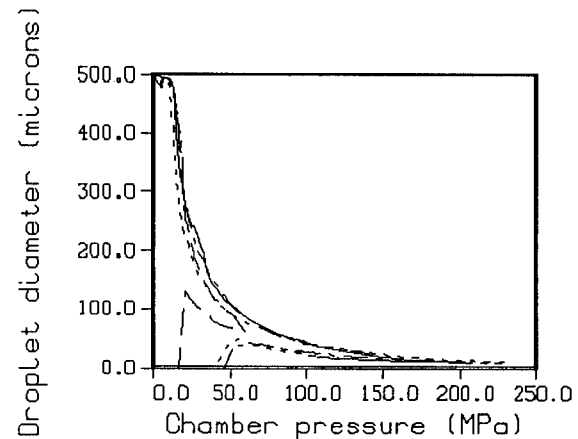
Figure 12. Droplet Injection Model. Combustion
Chamber Droplet Diameter Profiles. 30mm (line).
105mm (dot). 155mm - 2 Liter Charge (dash).
155mm - 5 Liter Charge (dot-dash).

Although there are differences between the curves in Figure 12, the profiles of injected droplet diameters show more similarities than the previously derived profiles (Figure 6) of the mean diameter of droplets

in the chamber. The similarity between the curves in Figure 12 is related to the fact that all the guns considered have almost the same hydraulic difference (ratio of projected are on the combustion chamber side of the control piston to projected area on the liquid reservoir side). Hence, the guns have similar injection velocities as a function of chamber pressure. The liquid and gas density, viscosity and surface tension are nearly the same in all four cases. Thus, since the injected droplet diameter predicted by Eq. (2) depends most heavily on the injection velocity, the droplet profiles in Figure 12 are very similar. In general, the fact that the profiles in Figure 6 and the profiles in Figure 12 both result in accurate simulations indicates that the combustion chamber pressure is not highly sensitive to small changes in the droplet profile.

INTACT JET CORE

The various droplet models studied predict very small droplets under gun conditions, and the droplet diameters must be multiplied by a constant to obtain agreement with experiment. This raises the question of whether immediate droplet formation is the most appropriate physical model in the gun environment. It is expected that some liquid will be in the intact core of the injected liquid jet. In this section we assess whether the experimental results can be better explained by assuming that most of the liquid accumulation is due to liquid in the intact core of the jet.

Mayer[12] developed a formula for the breakup time of the jet. His result can be written in the form

$$t_b = 0.2356 \, D_h \, V^{-4/3} \, \sigma_L^{1/3} \, \mu_L^{-1/3} \, \rho_L^{2/3} \, \rho_G^{-2/3} \, , \qquad (3)$$

where D_h is the hydraulic diameter of the annular injector. The length of the intact core can be approximated by multiplying the breakup time t_b times the injection velocity V (ignoring drag). The volume of the intact core is then approximately

$$Vol = 0.5 \, V \, t_b \, A_v \, , \qquad (4)$$

where A_v is the vent area of the injector.

Wolfe and Andersen[11] developed a similar formula

$$t_b = D_h \, / \, [\, (A^2 + B \, P)^{1/2} - A] \, , \qquad (5)$$

where

$$A = 16 \, \mu_L \, / \, D_h \, \rho_L$$

$$B = 2 \, / \, \rho_L$$

$$P = 0.5 \, \rho_G \, V^2 - 2 \, \sigma_L \, / \, D_h \ .$$

The authors note that if the viscous and surface tension forces are negligible, this reduces to

$$t_b = D_h \, V^{-1} \, \rho_L^{1/2} \, \rho_G^{-1/2} \ . \tag{6}$$

Chatwani and Bracco[13] report a breakup length formula based on the theory of Taylor. The breakup length is

$$x_b = (1.75 \, / \, f_m \,) \, D_h \, \rho_L^{1/2} \, \rho_G^{-1/2} \ , \tag{7}$$

where f_m is a numerically computed quantity less than one and the constant, 1.75, was chosen based on comparison with experiment. This is in fact identical with the simplified Wolfe and Andersen formula, except for the $1.75/f_m$ term.

All three of the formulas were implemented in a small post-processor code. The parameter values are read in from the output of a gun code simulation. Figure 13 shows the results for a 30mm firing. The solid line is the liquid accumulation from a simulation of the 30mm Concept VIC gun (dotted line in Figure 7). The dotted line is the accumulation resulting from the application of the Taylor formula Eq. (7). The dashed line gives the Wolfe and Andersen result, Eq. (5), multiplied by ten, and the dotted-dashed line the Mayer result, Eq. (3), multiplied by a hundred. The results were multiplied by a constant to obtain the proper order of magnitude for the accumulation.

Figure 13. Liquid Accumulation. 30mm Gun Simulation (line). Taylor (dot). Wolfe and Andersen times 10 (dash). Mayer times 100 (dot-dash).
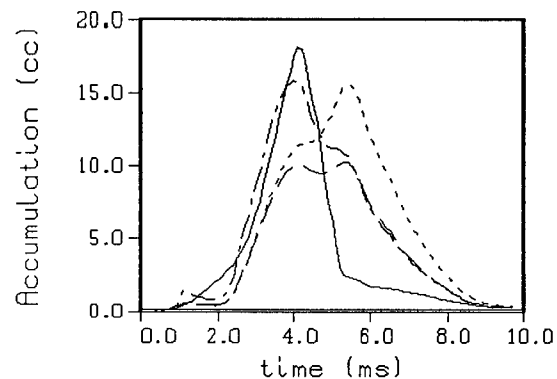
The computation of the intact jet length is known to be only approximate. A comparison of the predicted accumulation in Figure 13 emphasizes the scatter in the three formulas. According to the Mayer

formula, Eq. (3), the liquid core is negligible. According to the Wolfe and Andersen formula, Eq. (5), the liquid core is at times a noticeable amount of the liquid accumulation. According to the Taylor formula, Eq. (7), the liquid core does in fact contain the same order of magnitude of liquid as required by the simulation.

It is more revealing to look at the shape of the curves. In the simulation, the liquid accumulation falls rapidly around 4 ms. However, all the models show the mass in the intact core staying high, and dropping off at a later time. Unlike the injected droplet diameter model based on instantaneous formation of droplets, the combustion models based on a large liquid core do not match the experimental data, even with scaling.

Figure 14 shows the results of an application of formulas in Eqs. (3, 5 and 7) for a 155mm firing (5 liter charge - see Figure 10). The difference in shape is even more dramatic. Also, totally different scaling factors than those used in the 30mm simulation were required to match roughly the magnitude of the liquid accumulation. Thus, an assumption that most of the liquid accumulation is in the intact liquid core does not match the experimental data using any of these three formulas. These formulas are based on questionable assumptions, and experimental studies of the intact cores of jets have only been performed at low pressures. Thus, the above study does not eliminate the possibility that there is a substantial amount of liquid in the intact core of the jet. However, at present, we are not able to model this phenomena, and it is not included in our simulations.

Figure 14. Liquid Accumulation: 155mm Gun - 5 liter Charge. Simulation (line). Taylor divided by 5 (dot). Wolfe and Andersen times 3 (dash). Mayer times 25 (dot-dash).
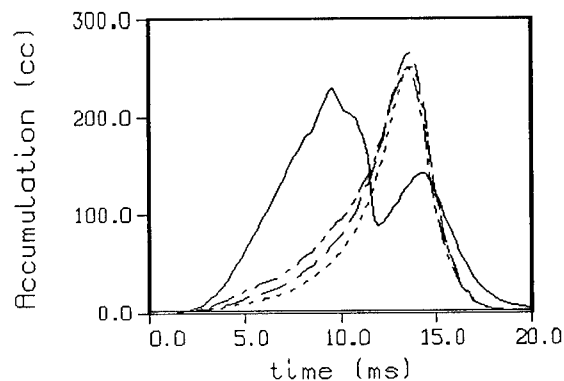
There remains the question of why the injected droplet diameter formula in Eq. (2) is applicable, with scaling, to the gun case. The droplet diameter formulas in Eqs. (1-2) are for droplets initially stripped off the surface of the jet and predict very small droplets. In our application we have biased the droplet diameter by ignoring the intact jet core. In addition, the sheet thickness is larger than that considered in the experiments, the gas is more dense, and the jet is

axially confined. Yet, apparently, the functional relationship between dominant parameters determining injected droplet diameter in Eq. (2) is approximately descriptive in the four cases considered. The fact that the 30mm, 105mm and 155mm cases are related in design and have similar injection velocity histories together with the strong dependence of droplet diameter in Eq. (2) on injection velocity may explain the similarity in injected droplet diameter profiles.

SENSITIVE TIME LAG MODEL

The droplet model is not completely satisfactory, since the theoretical correlation must be scaled to obtain good agreement with the experimental data. An alternate approach is developed, based on a simpler combustion model.

Liquid propellant rockets are similar in structure to liquid propellant guns. Crocco[14] has developed a time lag model for combustion in a rocket. As discussed above, there are many phenomena occurring in the injection and combustion process. In Crocco's theory, all of these phenomena are lumped into a single time lag. That is, the liquid propellant is assumed to be injected and, after some specific time lag τ, to instantaneously release all of its energy and form final products.

If the time lag τ varies with time according to some known relation, Crocco derived the governing equation

$$m_r(t + \tau) \; [1 + \partial\tau/\partial t] = m_j(t) \; , \tag{8}$$

where m_j is the injection rate (g/s) and m_r is the energy release rate (g/s). The derivative of the time lag τ is evaluated at the time t. If the time lag is constant, this reduces to $m_r(t + \tau) = m_j(t)$. If the time lag is increasing, the energy release is spread out over a larger time interval than the injection, i.e., the energy release rate is smaller than the injection rate. If the time lag is decreasing, the energy release occurs over a smaller time period than the injection process, and the energy release rate is larger than the injection rate.

In its final form, the time lag is assumed to be sensitive to the conditions in the chamber. That is, the time lag depends on the environment of the injected liquid between the time of injection and the time of energy release. Crocco assumed that the time lag is sensitive only to the chamber pressure. When the integral of pressure over time reaches some predetermined delay constant, the propellant energy is released. All other phenomena that affect the jet breakup and combustion are assumed implicitly to correlate with chamber pressure. As the pressure increases, the time lag becomes smaller, which is qualitatively correct.

RESULTS - TIME LAG MODEL

As before, the model is first applied to the 30mm case. A value of 30 MPa-ms for the delay constant gives the best results (Figure 7, dashed line). The sensitive time lag simulation is very similar to the simulation using the droplet combustion injection model (dotted line). The interior ballistic model is applied to the other three gun cases, using 30 MPa-ms for the delay constant. The overall agreement between simulated and experimental pressure versus time profiles is equivalent to that obtained using the droplet combustion model (Figures 8-10, dashed line). The agreement in muzzle velocity is also good (Table 1). The largest disagreement in the maximum chamber pressures is about four percent (30mm case).

The results are very good considering the simplicity of the model, i.e., droplet size and propellant combustion rates no longer enter into the combustion model. The results are similar for the other seven cases of 30mm, 105mm, and 155mm data (not shown).

The agreement between the predicted pressure profiles from the injection model and the sensitive time lag model is very good. This is partly due to the nature of the gun configurations. The droplet size in the droplet injection model primarily depends on the velocity of the injected liquid. This velocity in turn depends on the pressure difference between the reservoir and the combustion chamber. The reservoir pressure will always be close to the chamber pressure times the hydraulic difference. So indirectly, the droplet size is most sensitive to the chamber pressure, which is also the parameter that affects the sensitive time lag model.

SUMMARY

A new droplet model has been introduced into the regenerative liquid propellant gun code. Unlike the previous droplet profile model, there is a single adjustable parameter. All of the experimental Concept VIC data is represented quite well using the same value of the adjustable parameter, which was fixed in the simulation of the 30mm test data.

Models of an intact jet core have also been studied. These models do not appear to be applicable to the gun cases considered. In general, intact jet core models predict an incorrect amount of accumulation, and accumulation occurring later in the ballistic cycle, than the experimental data.

An even simpler combustion model is introduced, in which the energy in the propellant is released after the cumulative pressure times time reaches a predetermined value. The results are remarkably similar to the new droplet model.

The processes in the combustion chamber are very complicated. None of the above models embodies the details of the injection, breakup, and combustion processes. However, for the 30mm, 105mm and 155mm firings

studied, the simple models are surprisingly accurate. Of interest is whether these models will extrapolate to firings of larger charge sizes planned for the 155mm RLPG during the coming year.

REFERENCES

1. Coffee, T.P., Wren, G.P., and Morrison, W.F., "A Comparison between Experiment and Simulation for Concept VIC Regenerative Liquid Propellant Guns. I. 30mm," BRL-TR-3072, December 1989.

2. Coffee, T.P., Wren, G.P., and Morrison, W.F., "A Comparison between Experiment and Simulation for Concept VIC Regenerative Liquid Propellant Guns. II. 105mm," BRL report, to be published.

3. Wren, G.P., Coffee, T.P., and Morrison, W.F., "A Comparison between Experiment and Simulation for Concept VIC Regenerative Liquid Propellant Guns. III. 155mm," BRL report, to be published.

4. McBratney, W.F., "Windowed Chamber Investigation of the Burning Rate of Liquid Monopropellant for Guns," ARBRL-MR-03018, April 1980.

5. McBratney, W.F., "Burning Rate Data, LPG 1845," ARBRL-MR-03128, August 1981.

6. Coffee, T.P., "A Lumped Parameter Code for Regenerative Liquid Propellant Guns," BRL-TR-2703, December 1985.

7. Coffee, T.P., "An Updated Lumped Parameter Code for Regenerative Liquid Propellant In-Line Guns," BRL-TR-2974, December 1988.

8. Morrison, W.F. and Coffee, T.P., "A Modified Lagrange Pressure Gradient for the Regenerative Liquid Propellant Gun," BRL Technical Report No. BRL-TR-3073, January 1990.

9. Faeth, G.M., "Evaporation and Combustion of Sprays", Prog. Energy Combust. Sci., Vol. 9, pp. 1-76, 1983.

10. Weiss, M.A. and Worhsam, C.H., "Atomization in High Velocity Airstreams", ARS Journal 29, 252 (1959).

11. Wolfe, H.E. and Andersen, W.H., "Kinetics, Mechanism, and Resultant Droplet Sizes of the Aerodynamic Breakup of Liquid Drops", Aerojet Report No. 0395-04(18)SP, April 1964.

12. Mayer, E., "Theory of Liquid Atomization in High Velocity Gas Streams", ARS Journal 31, 1783 (1961).

13. Chatwani, A.U. and Bracco, F.V., "Computation of Dense Spray Jets", 3rd International Conference on Liquid Atomization and Spray Systems, Imperial College, London, July 1985.

14. Crocco, L., "Theoretical Studies on Liquid-Propellant Rocket Instability", Tenth Symposium (International) on Combustion, pp. 1101-1128, The Combustion Institute, 1965.

Electrical and Microstructural Characterization of Ion Implanted
Polycrystalline Silicon at High and Low Temperatures (U)

*Melanie W. Cole, Ms.
U.S. Army ETDL, Ft. Monmouth, N.J. 07703

INTRODUCTION

Polycrystalline silicon is a material of great importance in microelectronic technology. It is used as gate contacts in metal oxide semiconductor devices, as the emitter in bipolar transistors and as an interconnecting medium between devices. Thin-film transistors fabricated in poly-silicon, are also important as these devices have strong potential in future 3-D integrated circuits, and fabrication of flat panel displays.[1,2]

Many of the electrical properties of polycrystalline silicon depend on grain size, and in general a large grain size is desirable.[1] Ion implantation is widely used to introduce dopants into poly-silicon, and enhance the grain size of polycrystalline silicon films on amorphous substrates.[3,4,5] The implantation causes the film to become amorphous and subsequent annealing results in larger grains than in the as-deposited layer. The use of higher ion implant beam currents is desirable from the processing point of view in order to achieve shorter implant times while still maintaining a minimum sheet resistance. An important consideration of using a higher beam current, with its resultant higher temperature, is the effect on the wafers electrical properties and microstructure. In order to better understand this phenomenon this work studied two processing situations; a room temperature As implant (230μA beam current) into poly-Si, and a 400°C As implant (500μA beam current). A secondary experiment consisted of a two step anneal using the low beam current (low temperature) implanted sample. Understanding the mechanisms involved between grain size enhancement and implant beam current/temperature will aid in the development of further refining the implantation-based processes involved in producing device quality materials.

EXPERIMENTAL

Poly-Si films, 220 nm in thickness, were deposited on 200 nm thermally grown SiO$_2$ over lying a Si wafer substrate using standard low pressure chemical vapor deposition (LPCVD) at 620°C. Two samples were implanted with Arsenic at 190 KeV with a dose of 2.0x10^{16} ions/cm^2. The only variable in the implant process was the ion implant beam current. One sample was implanted with a beam current of 230 μA (room temperature), while the second implant was performed at a higher beam current of 500 μA (400°C). Both samples were annealed in a Nitrogen atmosphere at 975°C for one hour. A secondary set of experiments were conducted involving only the pre-annealed low temperature (230 μA beam current) As implanted sample. A series of two step annealing experiments were performed by varying the first annealing temperature from 550°C to 650°C for 50 minutes while the second step remained constant at 975°C for 10 minutes.

A Philips 420-T transmission electron microscope (TEM) at 120 KeV was used to examine the poly-Si grain size. Thin foils were prepared in plan view back etched via HF:CH$_3$COOH:NHO$_3$ chemical solution. Secondary Ion Mass Spectrometry (SIMS) was used to achieve dopant profiles for both implants. Electrical characterization consisted of spreading resistance analysis (SRA) and carrier concentrations via electrical chemical methods (Polaron).

RESULTS AND DISCUSSION

The SIMS analysis revealed similar dopant profiles for both room temperature (230 μA beam current) and the 400°C high temperature (500 μA beam current) implants with an As concentration of 8.0x10^{20}/cm^3. However, the electrical characterization (Table 1) on the annealed samples revealed the high temperature implant to have a carrier concentration of 5.0x10^{19}/cm^3 with a sheet resistivity of 39 ohms/square, and the room temperature implant to have a carrier concentration of 1.0x10^{20}/cm^3 with a sheet resistivity of 24 ohms/square.

The microstructure of the pre-implanted poly-Si layer was observed to be composed of equiaxial small grains with an average grain size of 80 nm (Figure 1). After implantation at room temperature (230 μA beam current) complete amorphization of the poly-Si layer was obtained with an As dose of 2.0x10^{16}/cm^2. The amorphization results from the series of collisions of the implanted As ions with the atoms in the host material. Numerous small highly disordered regions (high concentrations of point defects such as interstitials and vacancies) form in the host and grow together to form a continuous amorphous layer. However, with the same dose implanted at 400°C (500 μA beam current) the poly-Si remained in the polycrystalline structure with only grain growth

	Sheet Resistivity	Carrier Concentration
Room Temp. Implant	$23.793\Omega/\square$	$1\times10^{20}/cm^3$
$400^\circ C$ Implant	$39.277\Omega/\square$	$5\times10^{19}/cm^3$

Table 1. Electrical characterization of As implanted poly-Si.

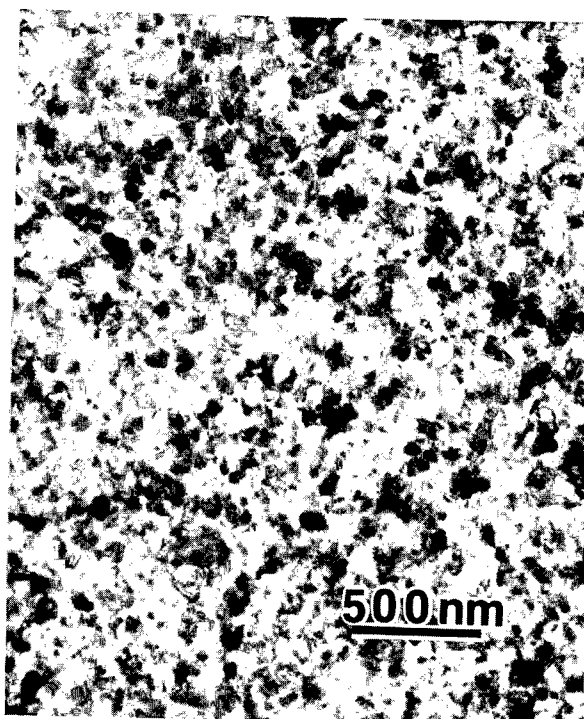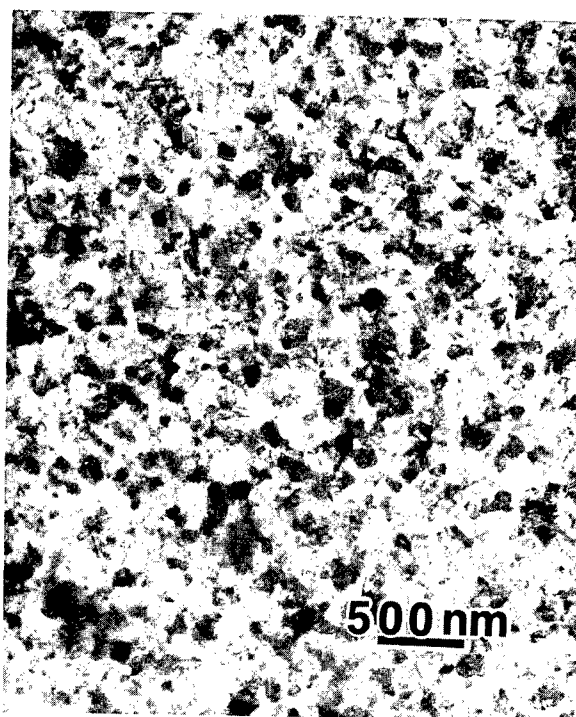Figure 1. Microsturcture of pre-implanted poly-Si sample.

Figure 2. Microstructure of pre-annealed 500µA As⁺ implanted sample.

from 80 nm to 140 nm (Figure 2). In this case the lattice damages by ion implantation have been annealed out dynamically during the implantation. This dynamic annealing results from the heating of the wafer as a result of the implantation process itself, i.e. from the higher beam current (500µA) with its associated higher substrate temperature (400°C). One possible proposed mechanism for this process suggests the generation of a very high temperature in the envelope immediately surrounding the path of the implanted ion. It is suggested that such localized thermal spikes may be conceived as achieving a higher amorphous to crystalline transformation rate as compared to that expected from the average equilibrated substrate temperature.[6] One may further argue that the formation of an amorphous zone is indeed temperature dependent. At high temperatures vacancies and interstitials have a higher mobility, thus these high temperatures permit annealing during the implant.

After annealing, the room temperature (230 µA beam current) implanted sample recrystallized and formed a polycrystalline structure of large grains with an average grain size of 1 µm (Figure 3). The driving force for the crystallization of amorphous Silicon films is the lower Gibbs Free Energy of the crystalline phase compared to the amorphous phase. The crystallization of amorphous Silicon, in solid state takes place by nucleation and growth where both processes are thermally activated.[7] The 400°C high temperature (500 µA beam current) implanted sample had no phase transformation, only grain growth to 260 nm (Figure 4). Figure 5 shows the annealed high temperature implant exhibited fewer defects within individual grains with respect to the room temperature implanted sample. Assuming the grains are three dimensionally equiaxial, the processes of annealing for these two implant conditions is schematically represented in Figure 6.

Calculated from the data of carrier concentration and dopant profile the electrical activation of As in the poly-Si layer is 12 % and 6 % respectively for the low and high temperature conditions. It is reasonably suggested that the electrical activation of dopant atoms is closely related to the microstructure of the annealed poly-Si layer. Grain boundaries are considered good sinks for impurities.[8] The smaller the grain size in the poly-Si layer, the larger the total area of grain boundary the poly-Si can have, there by increasing the possibility of As dopants being trapped in the grain boundaries and becoming electrically inactive. Thus, the smaller grain size of the 400° C high temperature (500 µA beam current) implant can be directly correlated to its lower electrical activity. This model is further supported by the qualitative Energy Dispersive X-ray Spectrometry (EDS) results presented in Figure 7. The EDS analysis with a 10 nm microprobe shows that there is a difference in As concentration between the grain boundary and mid-grain areas, with the grain boundary area showing the higher As concentration.
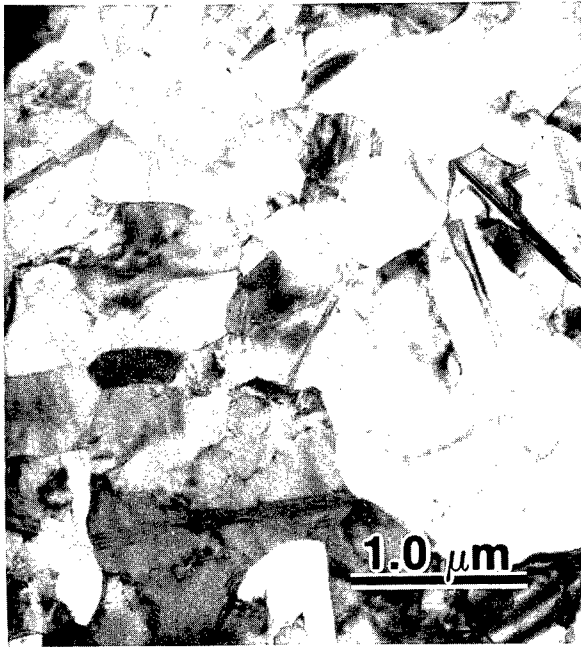
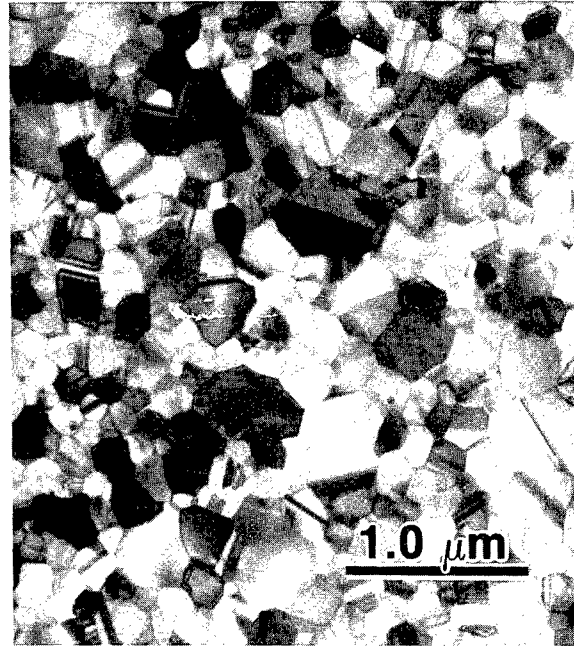Figure 3. Microstructure of annealed
230μA As⁺ implanted sample.

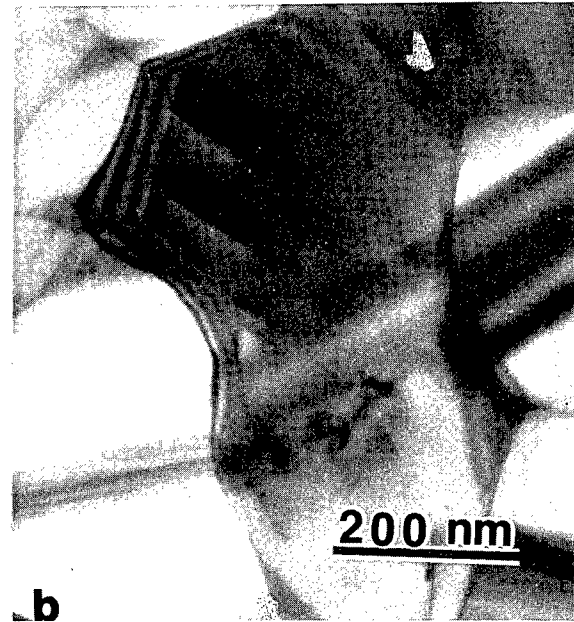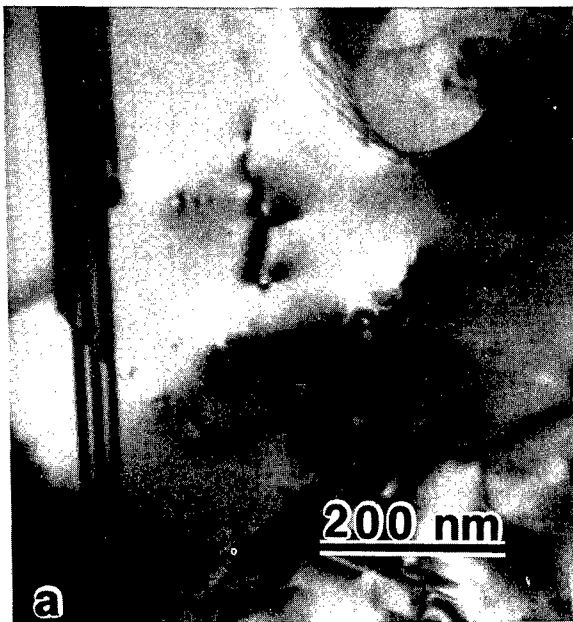Figure 4. Microstructure of
annealed 500μA As implant sample.

Figure 5. Defects within individual grains in a)230μA
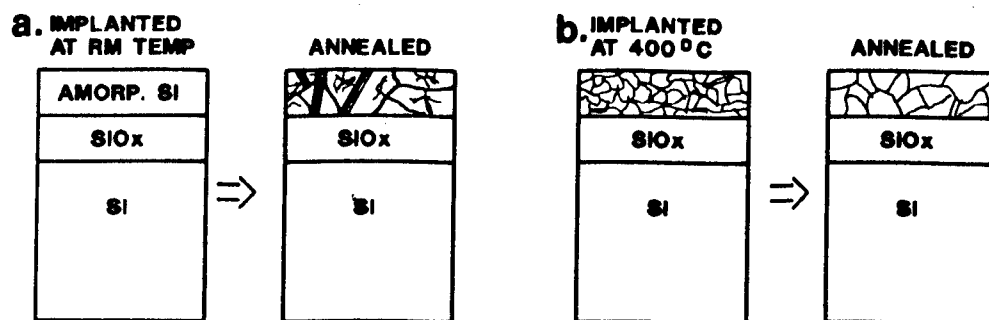implanted poly-Si, and in b)500μA implanted poly-Si.

Figure 6. Cross sectional schematic of the two different annealing processes.
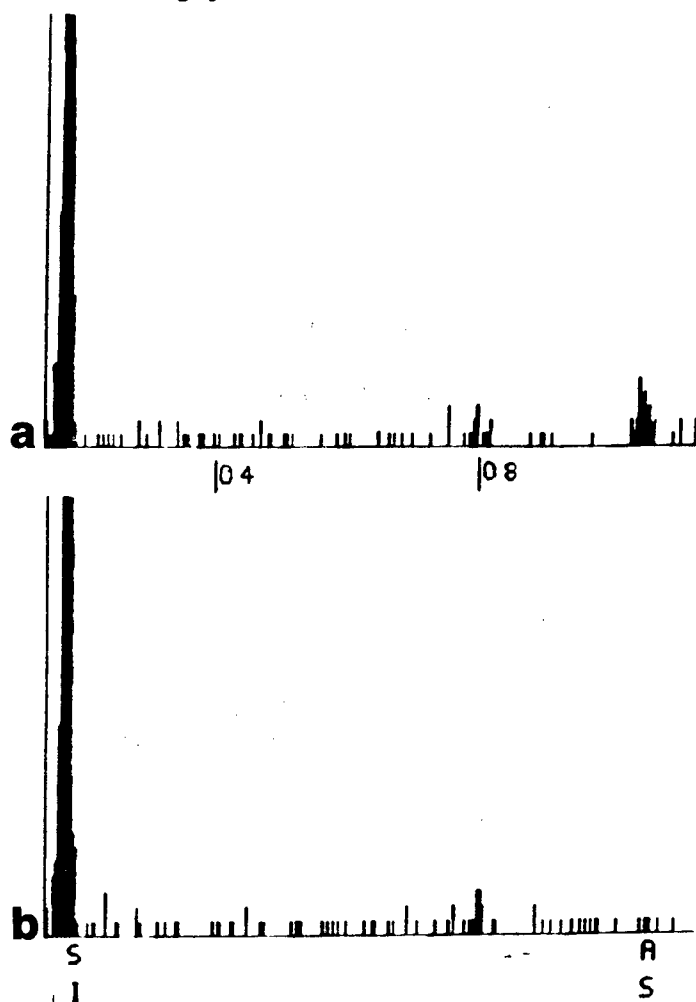

Figure 7. EDS results a) grain boundary area & b) mid-grain area.

Additional experiments were conducted with the intent to reduce the area of grain boundaries, i.e. to produce larger grains accompanied by improved electrical properties., The original one hour Nitrogen anneal at 975°C was replaced by a two step Nitrogen annealing process. The low temperature (230 µA implant beam current) was the only sample suitable for this experiment due to the amorphization of the poly-Si by the implantation. The first annealing step allows Silicon recrystallization at a low temperature (550°C to 650°C), this allows a reduced nucleation rate of poly-Si. The Silicon atoms are allowed to rearrange themselves in an ordered situation. Previous workers have demonstrated the interdependence of grain size and annealing temperature.[1,8] It is suggested that a larger grain size will be achieved at lower annealing temperatures (500°C to 700°C) provided the activation energy for nucleation of new crystallites is larger than the activation energy for grain growth (a difference of 0.25 eV between the two activation energies).[1] The second annealing step occurs at a high temperature of 1000°C for a short period of time in order to activate dopants. Table 2 displays these annealing results. The most favorable anneal, 650°C for 50 minutes + 975° C for 10 minutes, resulted in a 28 % reduction in resistivity verses the original single step anneal at 975°C for one hour (Table 1). Figure 8 shows that this decrease in resistivity is accompanied by an increase in grain size to 1.8 um.

Results of this study shows the reduction in processing time via higher implant beam currents to be inversely proportional to material (device) quality. Thus, process time is reduced in this manner at the expense of material quality, i.e. inferior electrical properties. This degradation of electrical properties is directly related to the microstructure of the poly-Si after the implant and to the mechanism of the subsequent anneal. The implantation via the higher beam current allows grains or parts of grains to survive the implant, thus amorphization is replaced by dynamic annealing. The surviving grains act as seeds in the subsequent anneal, thus the annealing mechanism is only a small amount of grain growth, with the final grain size slightly larger than that of pre-implant. The complete amorphization of the poly-Si resulting from the low temperature 230 µA beam current implant results in an annealing mechanism of recrystallization and growth of spontaneous nucleated grains producing a much larger grain size. Modification of the annealing process, via a two step anneal, seems to be a more useful process modification, since it achieves improved electrical characteristics within the same processing time.

CONCLUSIONS

Implementation of higher ion implant beam currents in order to achieve shorter implant processing times has considerable drawbacks in terms of device material quality. Use of a high implant beam current, 500 µA, produces a poly-Si film with a lower

Original Sample		Two Step Anneal		
	975°C 1hr	550°C 50min 975°C 10min	600°C 50min 975°C 10min	650°C 50min 975°C 10min
Sheet Resistivity	23.793Ω/□	19.003Ω/□	18.737Ω/□	17.005Ω/□
Resistivity decrease		20%	22%	28%

Table 2. Electrical characterization of the two step anneals.
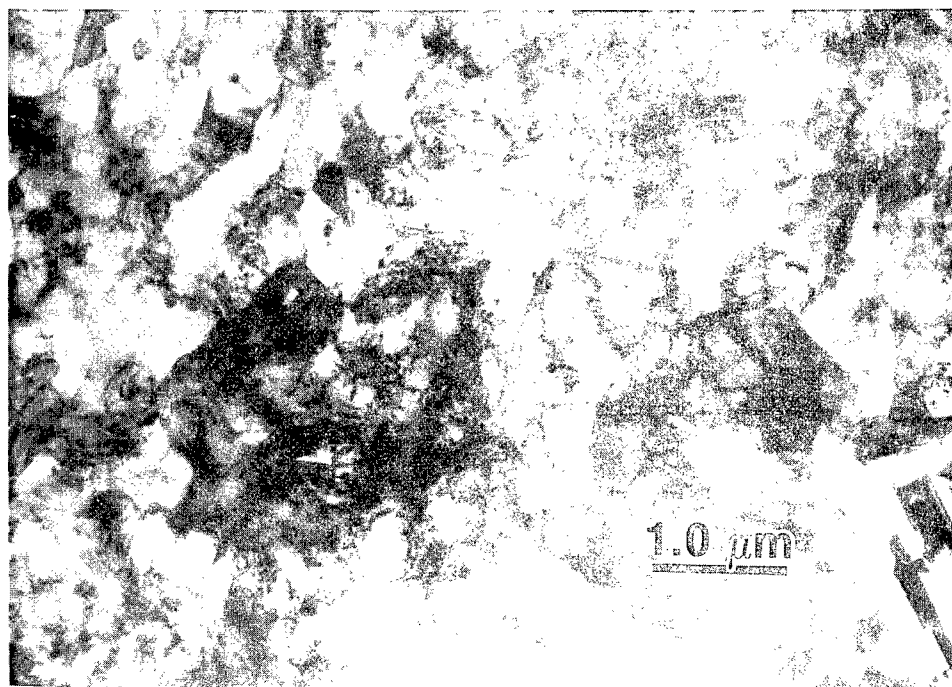
Figure 8. Microstructure resulting from the two step anneal.

electrical activity (dopant activation) than that of a lower beam current, 230 μA. This reduced electrical activity is directly related to both the microstructure after the ion implantation, and the mechanism of the subsequent anneal. The higher beam current produces smaller poly-Si grains resulting in a larger area of grain boundaries which serve as gettering sites for the As, and prevent the dopant from becoming substitutional and active. Use of the low implant beam current in combination with a two step annealing process has proved promising in achieving improved electrical properties of the poly-Si film.

REFERENCES

1. M.K. Hatalis and D.W. Greve, J. Appl. Phys. 63, 2260 (1988).

2. W.G. Hawkins, IEEE Trans. Electron. Devices, ED-33, 477 (1986).

3. R.B. Iverson and R. Reif, Mater. Res. Soc. Symp. Proc. 27, 543 (1984).

4. A.Bhattacharyya and K.N. Ritz, J. Electrochem. Soc., 131, 2143 (1984).

5. Y. Komen and I.W. Hall, J. Appl. Phys., 52, 6655 (1981).

6. S. Prussin, D.I. Margolise and R.N. Tauber, J. Appl. Phys., 54, 2324, (1983).

7. D.A. Porter and K.E. Easterling, Phase Transformations in Melals and Alloys, (Van Nostrand Reinhold, NY,) 263, (1981).

8. M.W. Cole, Y.C. Shih and C.F. Cook, Electron Micro. Soc. of Am. Proc., 46, 924, (1988).

Microstructural Characterization of Semiconductor
Materials as Related to Device Performance (U)

*Melanie W. Cole, Ms.
Mitra Dutta, Dr.
Peter Newman, Mr.
U.S. Army ETDL, Ft. Monmouth, N.J. 07703

INTRODUCTION

The military's need for high speed signal processing and real time information acquisition has encouraged GaAs semiconductor device research. Development and refinement of GaAs technology will serve to satisfy the projected defense requirements in the areas of directed energy weapons, laser and mm wave transmitters in target recognition and air defense systems. Device performance is directly influenced by the materials (GaAs) crystal quality, thus microstructural characterization of these high speed devices is essential in order to improve device quality via refined fabrication techniques. In this study the microstructure (interface configuration, atomic alignment, nature and extent of crystal defects) of hetero and regrown III-V interfaces and epilayers is characterized and correlated to device quality and fabrication techniques. Specifically, two device structures are emphasized: I. inverted-high electron mobility transistor (I-HEMT) and II. lateral quantum well arrays (QWA).

Past research has optimized the interface quality, hence mobilities (near 130,000 $cm^2/(Vs)$) for device structures such as the high electron mobility transistor, HEMT (MBE growth of Al Ga As on GaAs layer).[1] Recently it is of interest for certain device applications to have similar mobilities for heterojunctions in which GaAs has been epitaxially grown on $Al_x Ga_{1-x} As$ (0.25<X<0.30). This structure is referred to as the Inverted-HEMT (I-HEMT). The I-HEMT has the potential for exceptionally high speed devices since the gate electrode can be brought very close to the two dimensional electron gas (2-DEG). The interface configuration of the AlGaAs/GaAs heterointerface is an extremely important consideration since the quality of the interface directly influences the devices electronic properties such as mobilities.

The lateral quantum well array structure has applications which include picosecond switches, high frequency mm wave components, and solid state lasers. Suggested' techniques for fabrication requires molecular-beam epitaxy (MBE) regrowth on reactive ion etched (RIE) semiconductors.[2,3] To date, little work has been done on assessing the material quality of the etched surface and regrowth epilayer. Assessment of material damage, introduced by the RIE process is important, as crystal damage leads to deterioration of device performance.

I. INVERTED-HIGH ELECTRON MOBILITY TRANSISTOR (I-HEMT)

EXPERIMENTAL: I-HEMT

The I-HEMT structure was grown on semi-insulating (100) GaAs in a Varian Gen II molecular beam epitaxy (MBE) machine. The substrate was cleaned via standard chemical procedures prior to loading, and the native oxide was removed thermally in the growth chamber under suitable As_4 pressure. The reflection high energy electron diffraction (RHEED) pattern and intensity were monitored, and growth rates and alloy composition were determined through oscillation periods. The growth was maintained at 0.25 µm/hr with a substrate temperature (T_s) of 610°C except during Si doping. The T_s for the Si doped region was lowered to 550°C and the doping density ranged from 5.0×10^{17}/cc to 1.0×10^{18}/cc. The structure was grown with one monolayer (ML) of Ga interjected every 40 and 10 ML of AlGaAs growth (followed by a 10s growth stop) of the initial (i.e. buffer) and final (i.e. spacer) undoped layers.

Microstructural characterization was performed using a Philips 420T STEM at 120 KeV, and an ISI 002B TEM at 200 KeV. Samples were prepared in <110> cross sections via mechanically polishing followed by ion milling with Ar^+ beam at 4 KeV in a liquid nitrogen cooled stage. The mobilities were determined via Hall measurements in the Vander Pauw geometry using In/Sn contacts.

RESULTS AND DISCUSSION: I-HEMT

The growth of a high quality GaAs on AlGaAs structure (I-HEMT), is difficult due to the high degree of disorder (roughness and impurity segregation) at the heterointerface. The major difficulties which compromise growth planarity and uniformity of the critical interface are three fold;
1) The slower migration kinetics of Al with respect to Ga
2) The high reactivity of Al to the ambient impurities and

3) The outdiffusion of the silicon dopant along the growth front towards the inverted interface.[1]

In order to overcome these difficulties, a novel-MBE technique involving use of RHEED which optimized the interface growth conditions was employed.[4] RHEED studies were performed on AlGaAs surfaces at low substrate temperatures (T_s=500°- 550°C vs 650° -700°C) required for reduction of the Si outdiffusion. Since reduced T_s leads to lower Ga and Al surface migration rates, the GaAs growth rates were lowered from 1.0 um/hr to 0.25 µm/hr to compensate. A growth interruption was employed in order to permit recovery of the growth front morphology. Figure 1 demonstrates RHEED specular intensity behavior under such conditions and reveals how the growth interruption permits recovery of the growth front morphology to a smoother surface upon termination of growth. In order to reduce the surface riding characteristics of ambient impurities which may be exaggerated further during growth stops, and to reduce the time required for surface recovery, one monolayer (ML) of GaAs was grown periodically, followed by a growth stop. Finally to insure a high degree of smoothness at the final AlGaAs growth front, the T_s was raised (630°-650° C) once Si doping was completed at the lower temperature (500°-550°C). A schematic of this structure is presented in Figure 2.

The electrical characterization of the I-HEMT structure grown under the above conditions demonstrates a 77 K mobility of 91,000 cm^2/(Vs), and a mobility under light of 107,000 cm^2/(Vs). These values are considerable improvement over conventional I-HEMT mobilities which have generally been in the range of 20,000 cm^2/(Vs). These high mobilities suggest a smooth interface between the AlGaAs and GaAs epilayers. Evaluation of this growth front, heterointerface, is achieved via TEM.

Low magnification (80,000X) bright field TEM examination shows the inverted interface (GaAs on AlGaAs) to be sharply defined (Figure 3). Dark bands correspond to the GaAs layers and the lighter band to the AlGaAs layer due to the difference in scattering factors.[5,6] The location of the inverted interface is deduced from the intensity differences. Figure 4 shows the dark field image of the same region formed with a (002) Bragg diffracted beam. The (002) diffracted beam was used because the (002) plane is parallel with the heterointerface, thus information on interface quality might be included in the image. The (002) reflections are kinematically forbidden in elemental semiconductors since contributions from the two fcc sublattices exactly cancel out the (002) reciprocal space position. In the zinc-blend structure, (GaAs, AlGaAs) the two sublattices are occupied by different atomic species and the (002) reflections are kinematically allowed. Thus, in the zinc-blend compounds these reflections are sensitive to the chemical nature of the occupants of the sublattice.[7] The (002) dark field TEM micrograph allows utilization of the differences in diffracted intensities between the GaAs and AlGaAs layers. The differences arises from the values of the structure factors of the two layers. The structure factor for the GaAs layer is:
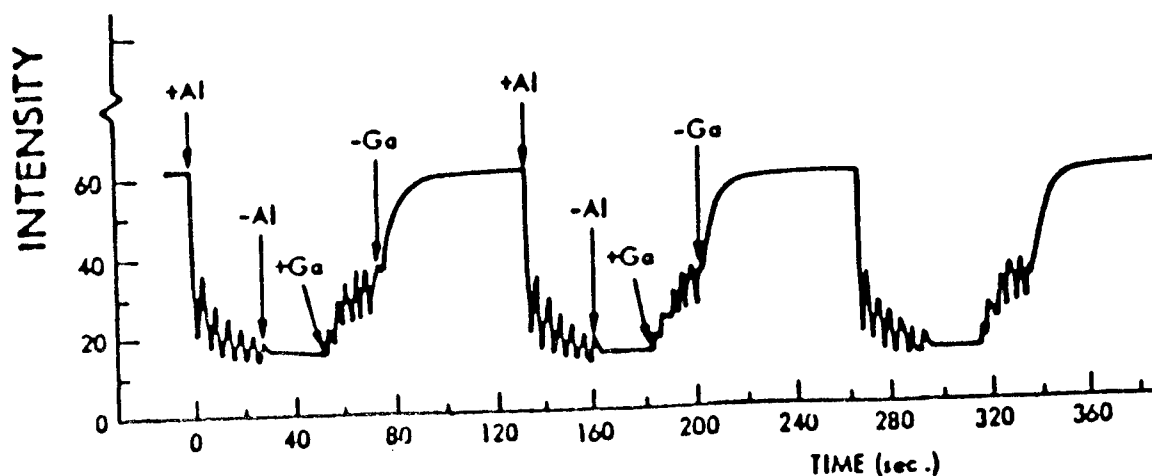
Figure 1. RHEED dynamic intensity demonstrates recovery of the AlGaAs growth front.
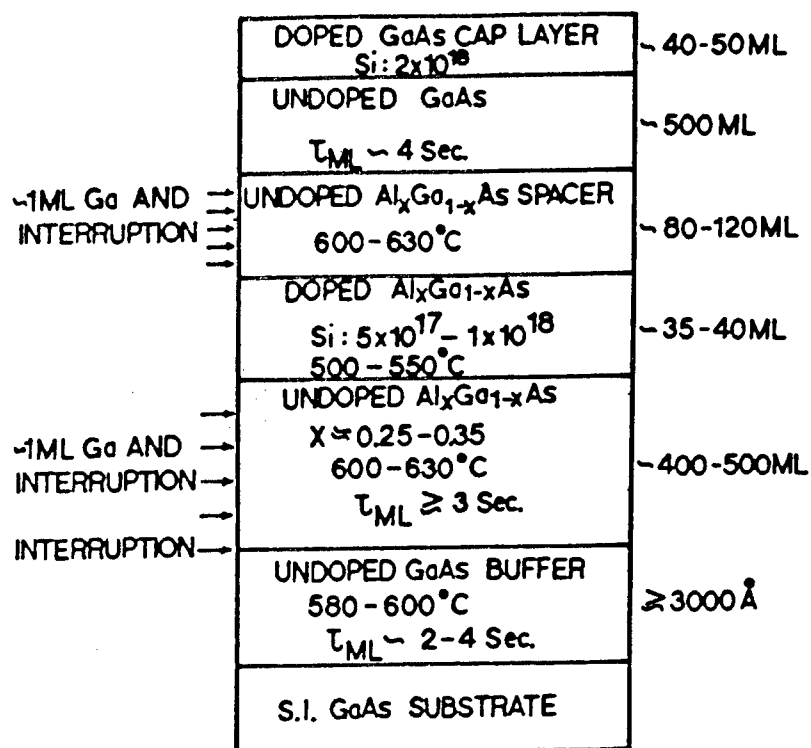
Figure 2. Schematic of I-HEMT structure, arrows mark 1 ML Ga and interruption.
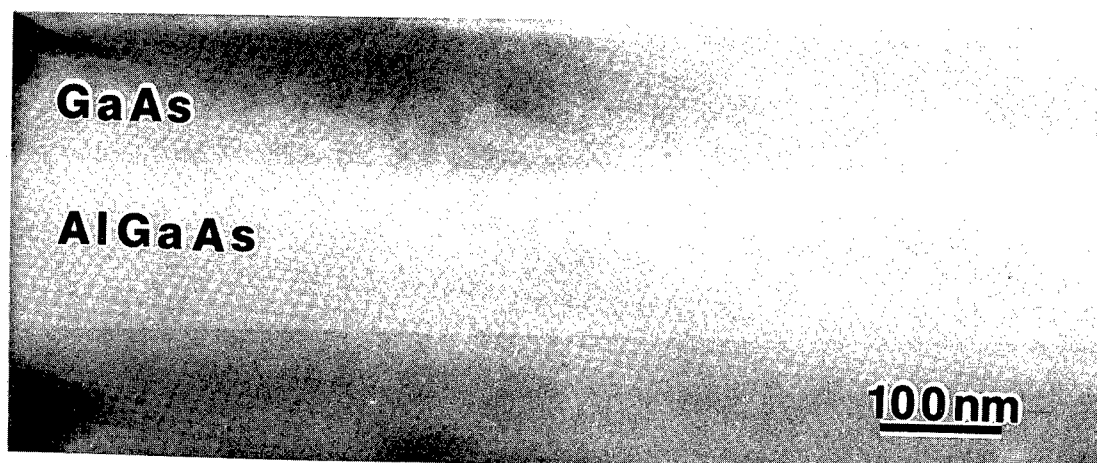
Figure 3. Bright field TEM micrograph showing inverted interface.
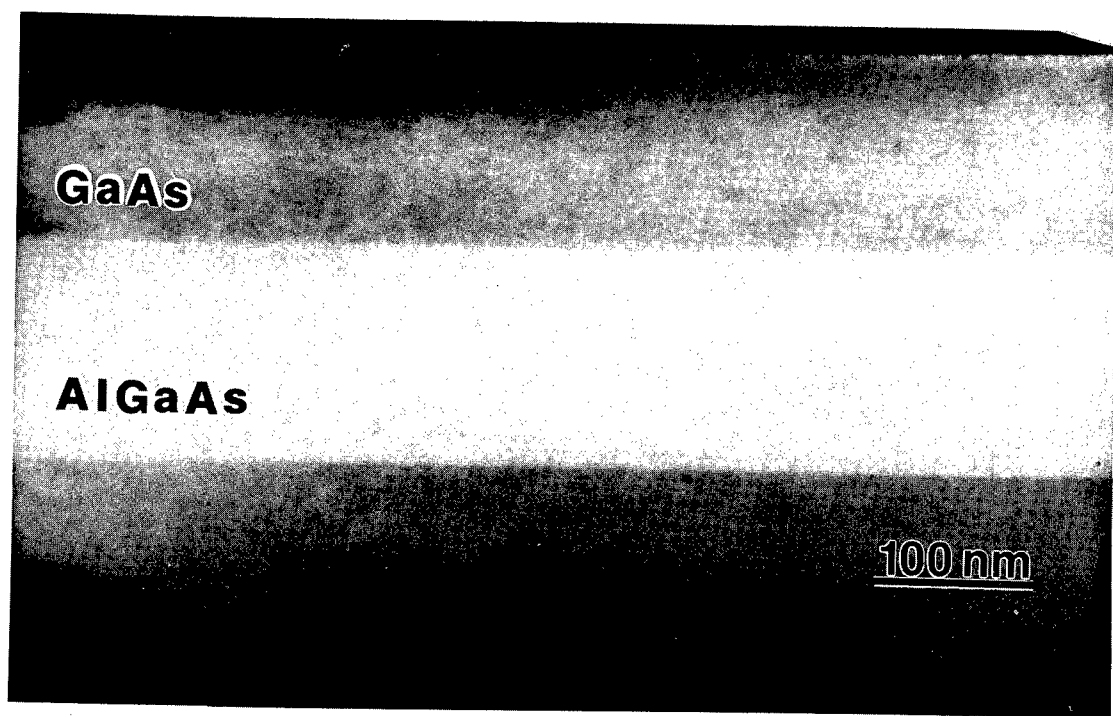
Figure 4. Dark field g=002 TEM micrograph showing the inverted interface.

$$F(002) = 4(f_{Ga} - f_{As}),$$

where the atomic scattering amplitudes of Ga and As are nearly identical ($f_{Ga} = f_{As}$). Thus, the (002) diffracted intensity is essentially zero and in the (002) dark field micrograph the GaAs layers appear as dark regions. However, the atomic scattering amplitude of Al (f_{Al}) differs significantly from that of As (f_{As}). For AlGaAs, the atoms occupying the Ga sites in the unit cell are either Ga or Al, and have an atomic scattering amplitude;

$$f(\theta) = (1-X)f_{Ga} + Xf_{Al}$$

In this case the structure factor for the (002) reflection is:

$$F(002) = 4[(1-X)f_{Ga} + Xf_{Al} - f_{As}] \approx 4X(f_{Al} - f_{Ga})$$

The diffracted intensity for a (002) dark field image:

$$I(002) = \frac{16\sin^2(\pi t s)}{\pi^2 k^2 \omega^2 s^2 \cos^2\theta}[f_{Al} - f_{Ga}]^2 x^2$$

where

ω ---unit cell volume
k ---wave vector of incident electrons
θ ---Bragg angle
t ---sample thickness
s ---deviation vector

Consequently the AlGaAs layer appears as bright regions.

In order to evaluate the abruptness of this heterointerface in more detail a cross sectional observation of the constituent lattice arrangement is necessary through implementation of high resolution electron microscopy (HREM). HREM is capable of directly and sensitively revealing the atomic configuration at the heterointerface, thus allowing evaluation of interface smoothness. Figure 5 shows the high resolution image of the inverted interface using [110] image orientation. The lattice images were constructed by an interference between the direct (000) beam, four {111} Bragg diffracted beams, and two {002} equivalent beams (Figure 6). The spacing between the
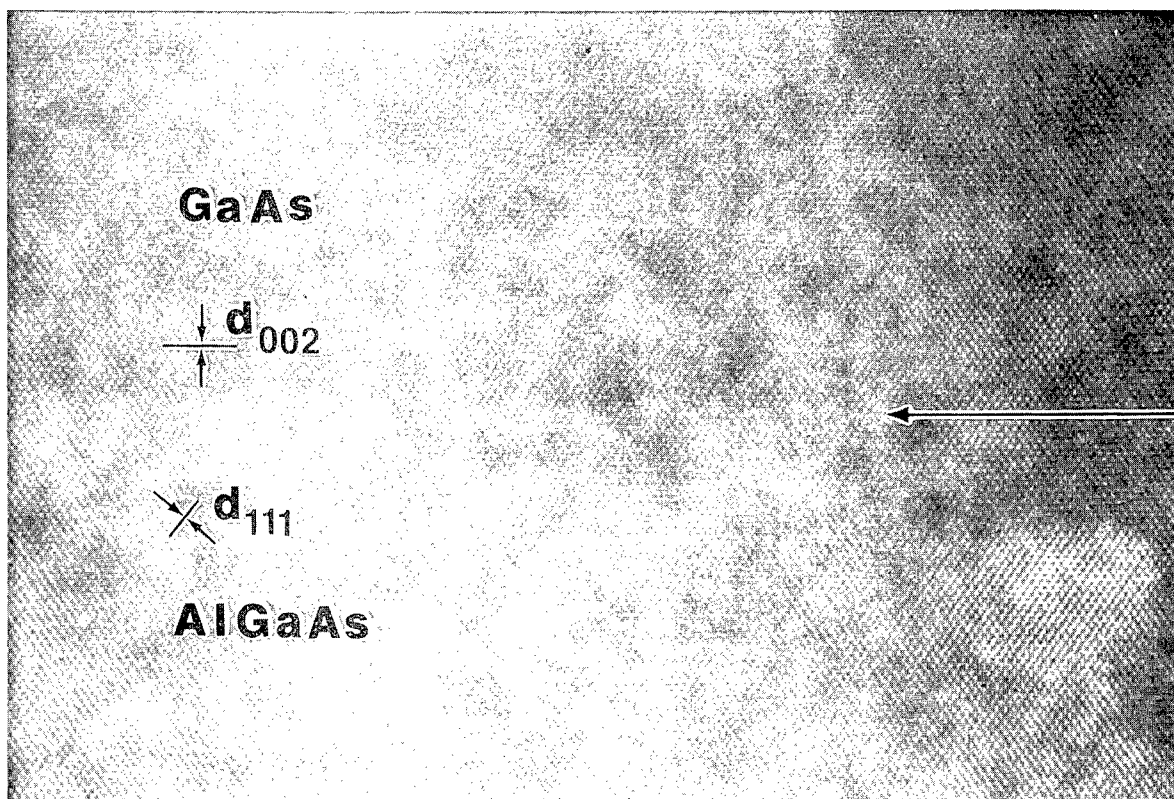
Figure 5. [110] lattice image of the inverted interface.
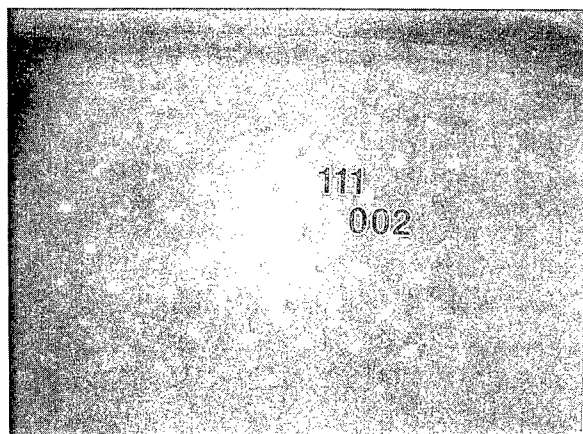
Figure 6. Diffracted beams used to form lattice image.

lattice images are 2.8Å for the (002) equivalent plane and 3.23Å for the (111) equivalent plane. The high resolution image shows that regularity in lattice alignment is well maintained and defect free across the AlGaAS/GaAs interface, and that the inverted interface is uniform to within a monolayer.

TEM microstructural characterization shows that the high mobilities of this structure are due to the atomically smooth interface. Thus, the MBE growth technique accomplishes its goal, smooth AlGaAs growth front, resulting in a high quality device structure.

II. QUANTUM WELL ARRAY (QWA): REACTIVE ION ETCHING (RIE)

EXPERIMENTAL: QWA & RIE

One micron of undoped GaAs was grown by MBE (Varion 360 MBE machine) on a semi-insulating GaAs substrate. The growth conditions were adjusted to achieve a growth rate of $1\mu m/hr$ with a substrate temperature of $620^\circ C$. Preparation of substrates has been described elsewhere.[8,9] Upon completion of this initial growth the wafer was cleaved into smaller pieces for exposure to different RIE conditions.

A parallel plate Plasma Therm PK 1250 reactive ion etch system at 13.56 MHz was used for all experiments. The system was evacuated to 1.0×10^{-6} Torr and then all process pressures were maintained at 10 mTorr with the cathode temperature of $^-40^\circ C$. Parameters which were not kept constant were the type of ion used for etching and etch power density. Specifically, HCl and Cl_2 reactive ion etches were performed at two power densities; $0.3W/cm^2$ and $0.6W/cm^2$. The samples were remounted and placed back in the MBE chamber to allow $1um$ of undoped GaAs to be grown on these etched surfaces. For comparison a sample was wet etched in a solution of $50H_2O:5H_3PO_4:H_2O_2$. Material characterization was of two types: microstructural via TEM and optical involving photoluminescence (PL) and Raman spectroscopy.

The RIE - regrowth samples were prepared in {110} cross sections for the TEM investigation. Samples were mechanically polished to $30\mu m$ then ion milled with an Ar^+ beam at 6 KeV in a liquid nitrogen cooled stage. Microscopy was performed using a Philips 420T STEM at 120 KeV and a JEOL 2010 TEM at 200 KeV. Microstructural evaluation was achieved via bright field, dark field, weak beam, and high resolution crystal imaging techniques.

PL spectra were taken at temperatures varying from 5K to 40K using an Argon laser's green line (514.5 nm) at a power density of $50 \ mW/cm^2$. The laser penetration

depth in the sample was 150 nm. The Raman experiments were undertaken using the 514.5 nm line from a Coherent Innova 90 Argon Ion Laser with a power of approximately 150 mW on the sample through a 100x microscope objective. A Z(XY)Z scattering geometry was used for all experiments. This configuration disallows the Transverse Optical (TO) mode by the selection rules based on crystal symmetry. Thus, the appearance of the TO phonons in the spectra are due only to crystal imperfections caused by the RIE conditions.[10]

RESULTS AND DISCUSSION: QWA & RIE

Reactive ion etching (RIE) is a preferred dry process step for fabrication of submicron III-V semiconductor device structures. This kinetically assisted chemical dry etch process offers several important advantages over other etching methods, namely etch anisotropy, etch rate control, selectively, accuracy, uniformity across the wafer and reproducibility.[11] As a result of these desirable characteristics, a number of semiconductor device designs involve RIE in their device fabrication process. Some of the more novel designs involve lateral quantum well arrays, whereby fabrication requires molecular beam epitaxial (MBE) regrowth on RIE semiconductors.[12,13]

A major drawback of RIE is the potential lattice damage introduced by energetic ion bombardment and contamination. This material damage may lead to deterioration of device performance. It is important to identify and understand this RIE induced damage in order to develop and refine the RIE process, i.e. minimize crystal damage and thus optimize device performance. This study identifies and analyzes the damage associated with various RIE parameters used for quantum well array fabrication.

Microstructural TEM evaluation shows all RIE - regrown samples exhibited lattice damage which emanates from the interface into the GaAs regrowth region. The extent and nature of the damage is dependent on the type of etchant and power density employed. The worse case was that of HCl at 0.6 W/cm^2. Material damage was extensive throughout the 1um regrowth region with defect density of > 1.0×10^{11}/cm^2. The presence of large microtwins, averaging 85 nm in width on {111} planes were observed to dominate the regrowth region. When the power density was lowered (0.3W/cm^2), while keeping the etchant constant (HCl) an appreciable decrease in crystal defects was observed. Figure 7 shows the material quality of the regrowth region for the 0.3W/cm^2 etch. There was a high concentration of material damage in the first 250 nm of the regrowth region. The defect density is ~1.3×10^{11}/cm^2 consisting of dislocation tangles and stacking faults and a few narrow microtwins averaging 50 nm in width. Material quality of the 0.3W/cm^2 Cl$_2$ RIE is superior to that of the HCl etchs. Figure 8 is a dark field micrograph of the GaAs - regrowth region. Lattice damage consists
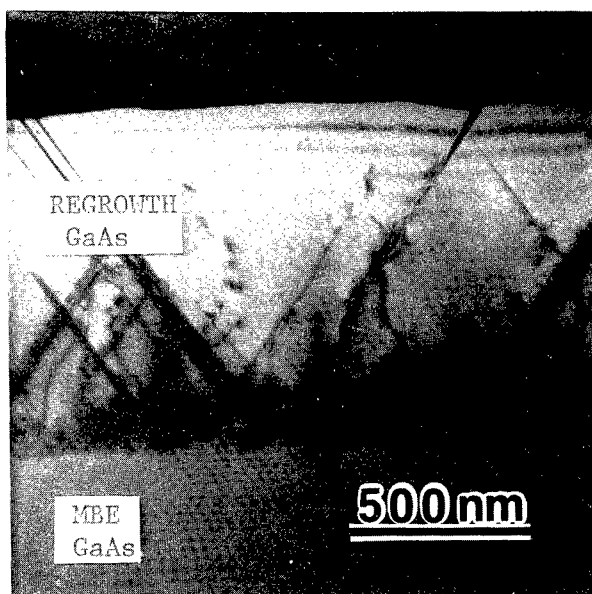
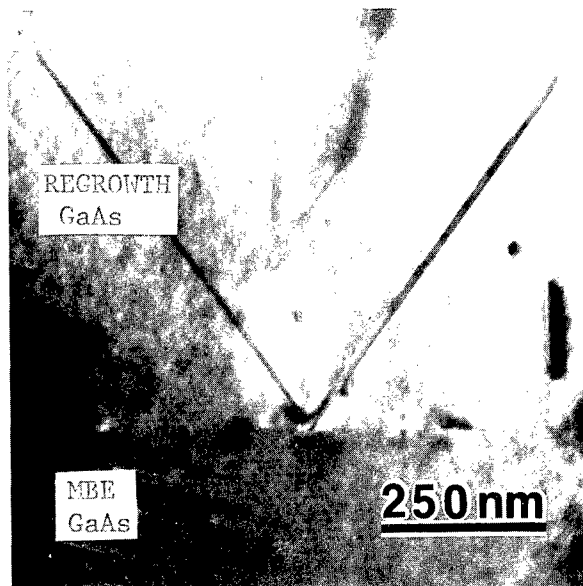Figure 7.   Dark field micrograph
of $0.3W/cm^2$ HCl RIE.

Figure 8.   Dark field micrograpn
of $0.3W/cm^2$ Cl_2 RIE.

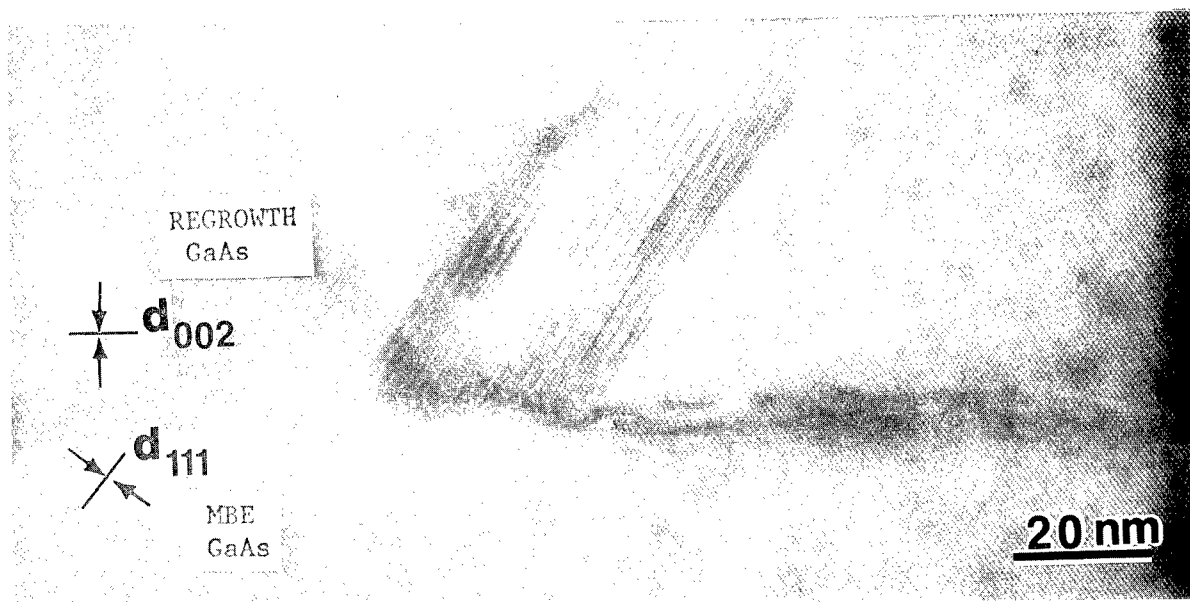Figure 9.   [110] crystal image of $0.3W/cm^2$ Cl_2 RIE.

358

primarily of dislocations and a few narrow ~18 nm {111} microtwins and stacking faults with a defect density two orders of magnitude less than that of the HCl etch at a similar power level. The extent of disorder into the regrowth region is ~ 150 nm. Figure 9 shows a [110] High Resolution Crystal Image of the Cl_2 RIE surface regrowth region. The {002} plane is parallel to the RIE surface. The {002} plane spacing is 2.83 Å and the {111} plane spacing is 3.26 Å. Planar defects along {111} planes show lattice shifts as stacking faults. The interface roughness is on the order of 6.5 nm. The defect density of the Cl_2 RIE sample is comparable to that of the wet etch.

Additional characterization of the epilayer crystal quality is achieved via PL and Raman spectroscopy. The PL data of the Cl_2 etched sample at a temperature of 5.4K revealed a strong peak at 1.5153 eV due to the exciton recombination and a smaller peak due to Carbon (1.4945 eV). In Figure 10 a comparison of exciton intensities of the HCl, Cl_2, and wet etchs vs temperature is presented. It is easily seen that the sample etched with Cl_2 has the highest intensity, thus the best crystal quality. The PL data mirrors the TEM data quite well, further illustrating the undesirable effects of the HCl RIE. Figure 11 compares the Raman spectra of the HCl and Cl_2 RIE at $0.3 W/cm^2$. TO scattering is ascribed to the presence of defects which destroy the crystal symmetry hence lead to a violation of the selection rule. Twinning and accompanying defects give rise to internal strain resulting in the appearance of the forbidden TO mode.[10,14] The intensity ratio of TO to LO can be used to qualitatively assess the crystallographic quality of the material.[15] A simple comparison of the TO/LO intensity ratio for the HCl and Cl_2 etchs shows the HCl to be larger which is indictive of the high density of twins and stacking fault defects revealed by TEM. Both optical and microstructural results confirm the inferior quality of the HCl vs Cl_2 etched material at equivalent powers.

Mu et al. has shown that RIE's which contain Hydrogen (CF_4/H_2) result in more structural damage to Silicon surfaces vs non Hydrogen plasmas (CF_4).[16] Similar studies of Silicon exposed to a Hydrogen containing RIE plasma attributes the crystal damage to be caused by the physical bombardment of the surface by Hydrogen ions.[17] The work of Jeng and Oehrlein has also realized that Hydrogen plays an important role in the formation of RIE related damage in Silicon. They suggest that light ion bombardment (H^+) causes the accumulation of point defects and point defect clusters similar to the implantation process. During the RIE process the cathode temperature is ~30-40°C. It is suggested that this temperature is enough to mobilize the point defects thus enabling them to recombine or diffuse to other clusters in order to form larger and stable defect clusters like stacking faults etc.[18] Our work seems to fit the above model well, as the Hydrogen containing etch shows more structural damage than the non Hydrogen etch (Cl_2). It is believed that the H^+ ions create more structural damage than Cl_2 ions because they move with a greater velocity and penetrate deeper into the crystal.[8,9,10]
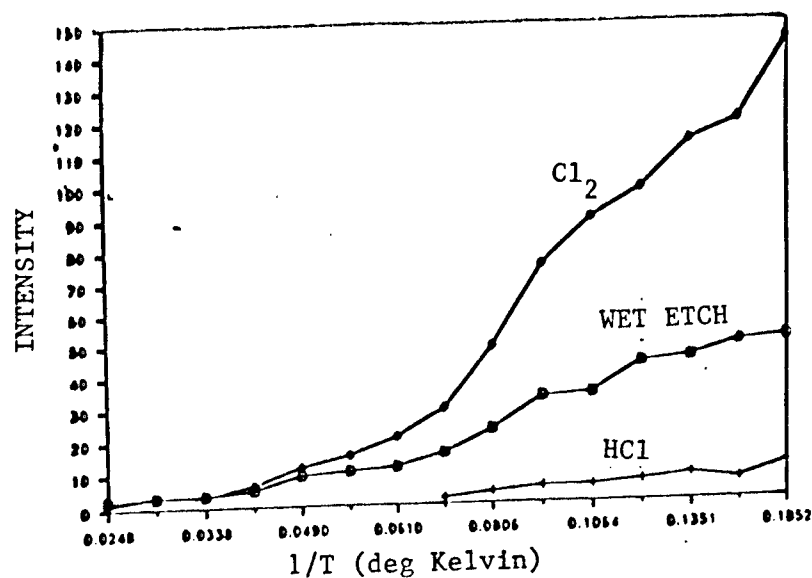
Figure 10. Comparison of exciton intensities of HCl, Cl_2, and wet etch vs temperature.
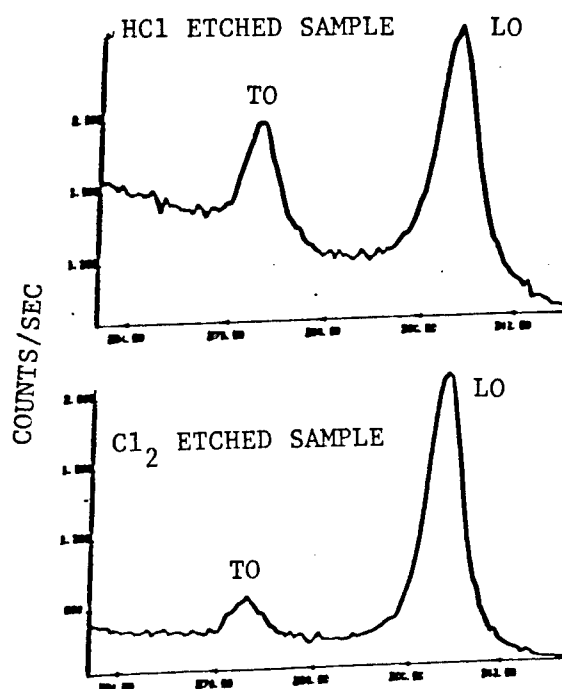
Figure 11. Raman spectra of HCl & Cl_2 at $0.3W/cm^2$.

Our results have shown that TEM, PL, and Raman spectroscopy are powerful tools for evaluating the damage induced by RIE processing. These characterization methods' have revealed that RIE using Cl_2 results in less material damage than that of HCl. The role of Hydrogen in the plasma is directly correlated to the high density of crystal defects observed in the GaAs regrowth region.

CONCLUSIONS

Results of this work clearly demonstrates that microstructural characterization of electronic materials is necessary to assess material damage and interface configuration, as the crystal quality is directly correlated to device performance. Characterization of the I-HEMT structure demonstrated that the interface smoothness was responsible for the exceptionally high mobility measurements. TEM characterization of this critical AlGaAs/GaAs heterointerface confirmed growth front smoothness via bright field, dark field, and lattice imaging techniques. The HREM images clearly revealed a planar interface showing well maintained lattice alignment with no local disorder. Optical and microstructural characterization of reactive ion etched-regrown GaAs (a method employed for QWA fabrication) showed the material quality to be inferior for the HCl etch versus the Cl_2 etch. It is suggested that the velocity of the H^+ ions combined with the elevated temperature of the RIE cathode are responsible for the increased damage seen in the HCl etch-regrown GaAs.

REFERENCES

1. N.M. Cho, D.J. Kim, A. Madhukar, P.G. Newman, D.D. Smith, T. Aucoin and G.J. Iafrate, Appl. Phys. Lett., 52, 2037 (1988).

2. D.A.B. Miller, J. Vac. Sci. Technol., B5, 313, (1987).

3. G.J. Iafrate, D.K. Ferry and R.K. Reich, Surface Sci., 113, 485, (1982).

4. P.G. Newman, N.M. Cho, D.J. Kim, A. Madhukar, D.D. Smith, T. Aucoin and G.J. Iafrate, J. Vac. Sci. Technol., B6, 1483, (1988).

5. M.W. Cole, J.L. Lehman, P.G. Newman, M. Dornath-Mohr, D.D. Smith, S. Guha and A Madhukar, Proc. of 47th EMSA Soc., 690, (1989).

6. Y. Suzuki and H. Okamoto, J. Appl. Phys., 58, 3456, (1985).

7. A. Ourmazd, Proc. of 45th EMSA Soc., 332, (1987).

8. L.S. Heath, D.D. Smith, M. Dutta, M.A. Taysing-Laura and T.P. Monaham, J. Electrochem. Soc., 136, 495, (1989).

9. J. Rossabi, M. Dutta, M.W. Cole and D.D. Smith, Proc. of 1989 SPIE meeting, (1989).

10. M.W. Cole, M. Dutta, J. Rossabi, D.D Smith and J.L. Lehman, Proc. of MRS Symp. G, (1989).

11. S.W. Pang, Electrochem. Soc., 133, 784, (1986).

12. D.A.B. Miller, J. Vac. Soc. Technol., B5, 313, (1987).

13. G.J. Iafrate, D.K. Ferry and R.K. Reich, Surface Sci., 113, 485, (1982).

14. P. Lao, W.C. Tang, A. Madhukar and P Chen, J. Appl. Phys., 65, 1676, (1989).

15. R.J. Matyi, W.M. Duncan, H. Shichigo and H.L. Tsai, Appl. Phys. Lett., 53, 2611, (1988).

16. X.C. Mu, S.J. Fonash, A. Rohatgi and J. Rieger, Appl. Phys. Lett., 48, 1147, (1986).

17. R.G. Frieser, F.J. Montillo, N.B. Zingerman, W.K. Chu and S. R. Mader, J. Electrochem. Soc., 130, 2237, (1983).

18. S.J. Jeng and G.S. Oehrlein, Appl. Phys. Lett., 50, 1912, (1987).

Theoretical and Experimental Aspects of a Short-
Scan Interferometer for Remote Chemical Detection

*Roger J. Combs, PhD; Robert T. Kroutil, PhD; and J.T. Ditillo, MS
U.S.Army Chemical Research and Development Center
Aberdeen Proving Ground, Maryland 21010-5423

MISSION:

The detection of toxic vapors by the use of a passive infrared sensor represents an important early warning military application. Current operational requirements demand the use of rugged sensors which operate in highly mobile battlefield environments. In this scenario, a vapor cloud must be detected while the InfraRed (IR) radiation background changes rapidly with sensors operating from helicopters or small Unmanned Aerial Vehicles (UAV). Traditional spectroscopic methods for toxic vapor detection with passive IR remote sensors are not reliable in mobile scenarios. Also, the traditional background subtraction technique that is used in fixed sight operation for extracting spectral features of interest assumes an initially clean environment. Modern signal processing methodology has been developed to eliminate problems associated with the background subtraction technique. These new methods include the use of Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) filters. These digital filters minimize unwanted background effects induced by the rapidly changing IR background radiation. Digital Signal Processors (DSP) implement digital filters to permit real-time detection of a target chemical vapor cloud. This recent capability allows the realization of a Short-Scan Interferometer (SSI) for remote chemical detection. This novel design results in a faster, compact, inexpensive and more rugged sensor that meets the challenge of the U.S. Army mobile chemical vapor detection mission.

PRINCIPLE OF DETECTOR OPERATION:

A passive remote sensing application is represented in Figure 1. In this scenario, a vapor cloud is present between the sensor and a terrain background. When the field-of-view of the sensor is totally filled, the radiance incident upon the sensor is a result of the emission characteristics in the 7 to 14 micron wavelength region of the background, the target vapor cloud and the intervening atmospheric gases. Assuming a homogeneous atmospheric and target cloud species, equation (1) gives the power (P) incident upon a passive sensor.[1]

$$P = (T_A T_B N_{BG} + (1 - T_A T_C) N_T) A_C \Omega_S \qquad (1)$$

C = CONCENTRATION
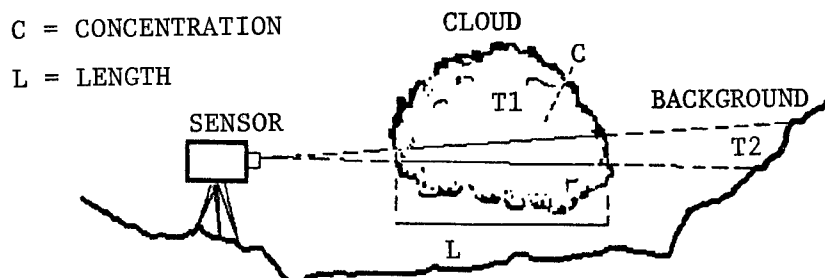
L = LENGTH

Figure 1. Operational parameters of (1) temperature difference, T2-T1, between cloud and terrain background, (2) cloud path length, and (3) cloud concentration define sensitivity of a passive IR sensor.

T_A is the atmospheric transmittance and T_B is the vapor cloud transmittance. N_{BG} is the background IR radiance and N_T is the radiance of a blackbody at ambient temperature. A_C is the collector area and Ω_S is the sensor solid angle of acceptance. The target cloud transmittance is given by equation (2).

$$T_C = \exp(-\alpha_T CL) \qquad (2)$$

α_T in units of m^2/mg is defined as the absorptivity coefficient of a particular target cloud species. C is the target cloud concentration in units of mg/m^3. L is the target cloud path length in units of meters.

One factor implicit in equation (1) is that the radiance values N_{BG} and N_T often differ by only a few percent. Vapor detection is not possible when $N_T = N_{BG}$, because $P = N_T A_C \Omega_S$ indicates that the incoming radiation contains no vapor cloud information. Thus, for vapor detection the occurrence of the condition $|N_{BG} - N_T| = \Delta N > 0$ is a requirement. The difference in radiance levels ΔN relates to a temperature difference between the background and the target vapor cloud. The relation for the ideal case is the Planck function. Therefore, the radiance difference for a particular sensitivity measurement corresponds to a temperature difference. Temperature differences between vapor clouds and backgrounds may range from 0.1 to 5.0 degrees centigrade. The use of a passive IR sensor for fixed sight applications is demonstrated with the U.S. Army XM21 passive remote chemical agent alarm. The XM21 is a conventional ruggedized Michelson interferometer that is used in the detection of chemical agent vapors. In ten years of XM21 field testing a case of the equivalence of background and target vapor cloud temperatures has never been documented.

SENSOR REQUIREMENTS:

The XM21, though not designed for mobile high vibration environments, is found to function in a limited capacity while mounted on a ground mobile or aerial vehicle. Previously used signal processing techniques of background subtraction fail to extract the target cloud spectral features in these mobile scenarios. Fixed sight processing requires collection of time domain data (i.e. interferogram) and subsequent conversion to the frequency domain

by a Fast Fourier Transform (FFT) for spectral analysis. The interferograms in Figure 2 contain spectral information for a range of frequencies; however, this spectral information is not distributed uniformly across the interferogram. For instance, data in the region of zero path difference for the interferometer is called the centerburst region and has significant low frequency background radiance information compared to the less dominant higher frequency information content. In contrast, data collected at the extreme mirror positions exhibits a substantial high frequency content but very little low frequency information. An intermediate region contains a large amount of spectral information for target vapor clouds interposed between the sensor and background radiation source. Low frequency background information remains superimposed on the spectral frequencies of interest.
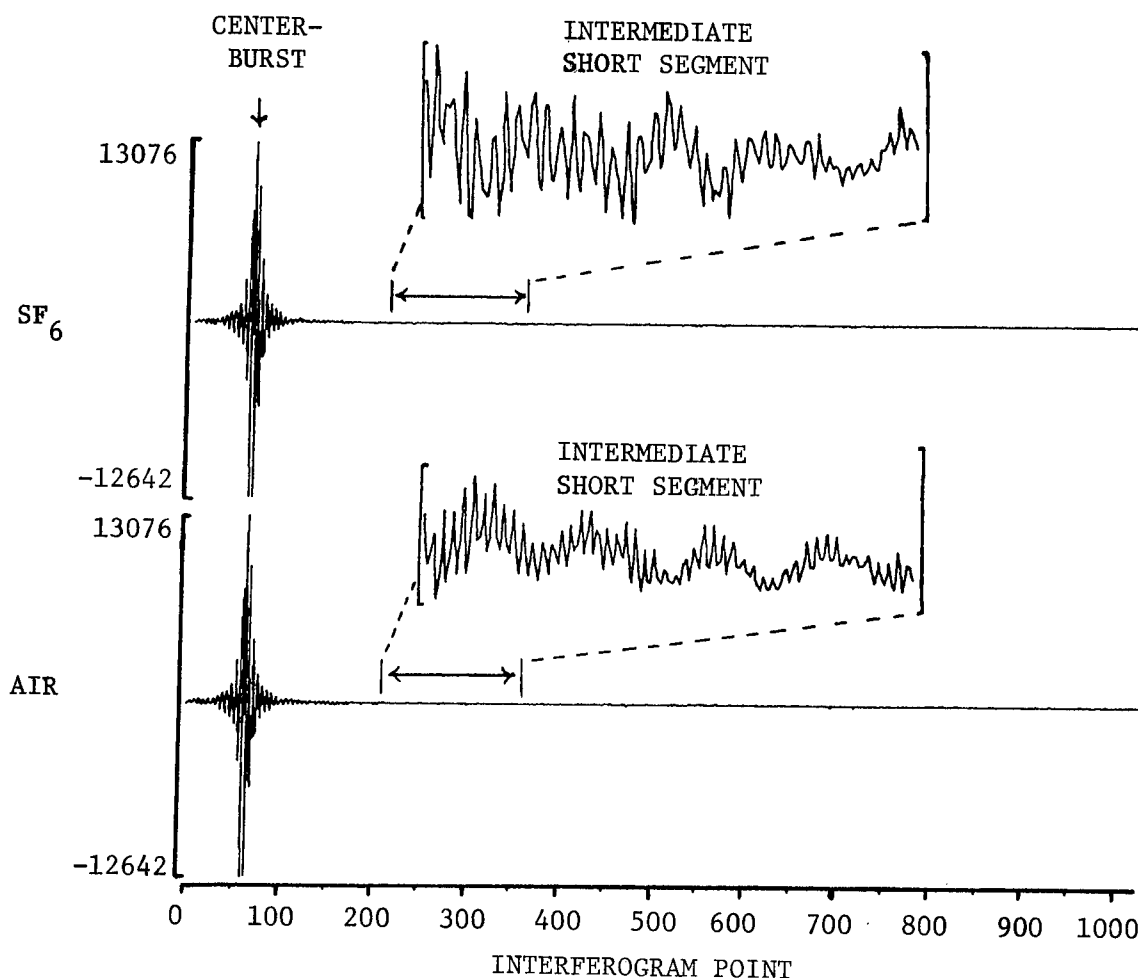
Figure 2. Short 150 point interferogram segment extracted from a full 1024 point interferogram.

The Moving Platform Algorithm (MPA) developed at the Chemical Research, Development and Engineering Center permits extraction of the target cloud spectral features from the rapidly varying background. This extraction method relies on the selection of an intermediate short segment from an entire full scan interferogram. The short interferogram segment is digitally filtered to remove the IR background signatures.[2] The background signatures are generally much wider in bandwidth than those of spectral features associated with the vapor cloud. Therefore, the reason for scanning the entire interferogram comes into question. Shorter scan lengths than those used by the XM21 results in greater insensitivity to vibration and a faster scan time. Higher scan rates essentially provide an interferogram that is collected instantaneously in a mobile environment rather than a smeared interferogram acquired at a slower scan rate. Signal processing of the short interferogram segment generated by a SSI is performed in the time domain as opposed to the previous computationally intense frequency domain processing for a full scan interferometer.

Modification of a full-scan interferometer into a SSI requires three considerations: (1) servo mirror control, (2) data acquisition electronics and (3) signal processing hardware. Proper signal sampling requires accurate mirror positioning that is provided by a servo mirror control circuit. The servo mirror control relies on fringe counting of the Helium Neon (HeNe) laser reference that is split into two signals with a quarter wave retarder. The two laser signals, which are in phase quadrature, permit determination of mirror direction as well as displacement from the interferogram centerburst. Short segment scanning depends on HeNe fringe counting from the centerburst using a set of proximity switches to maintain the correct mirror displacement.

Data acquisition of the short interferogram segment requires the appropriate data collection electronics. The data collection electronics consists of two components: (1) a bandpass analog filter and (2) an Analog to Digital Converter (ADC). Signal frequency specifications for these components depends on mirror modulation frequency and the spectral frequencies of interest. Equation (3) describes this relationship in terms of mirror velocity, v ; spectral feature of interest, $\bar{v}$; and resultant signal frequency, f.

$$2v\bar{v} = f \qquad (3)$$

Faster interferometer scan rates minimize interferogram smearing, but place stringent requirements on the data acquisition electronics. The analog filter must operate at a higher bandpass proportional to a faster mirror velocity as described in equation (3). The analog filter attenuates frequencies not of interest by eliminating the aliasing effects. Aliasing refers to erroneously identifying a high frequency noise component as a low frequency signal of interest. Removal of the aliased noise from a digitized signal is impossible with signal processing. Therefore, proper ADC sampling of the

interferometer signal demands bandwidth limiting the signal with a bandpass filter. The bandwidth limited signal generated from the SSI requires a lower ADC dynamic range or fewer ADC bits than a full scan interferometer. Rather than sixteen ADC bits for the full scan interferometer, the SSI uses from eight to ten bits dependent upon the selection of the short interferogram segment. The lower ADC resolution gives advantages in a reduction of the ADC conversion time and the sampling error. These advantages in the lower ADC resolution provide a means of capitalizing on a faster interferometer scan rate over the short interferogram segment.

The use of digital filtering to extract spectral features from a short interferogram segment demands high performance digital signal processing hardware. Digital Signal Processors (DSPs) due to their special architecture perform convolution calculations (i.e. repetitive sum-of-products computations) about ten to twenty times faster than a general purpose processor operating at the same clock speed. The faster DSP program execution times allow operation of the SSI in a real-time mode. All digital filtering calculations with this dedicated reduced instruction set computer are completed between the interferometer mirror scans. Realization of a self-contained, lightweight and mobile IR sensor requires the use of a dedicated reduced instruction set computer.

TIME DOMAIN PROCESSING:

Digital filters allow extraction of the spectral features associated with a target vapor cloud. A digital filter provides a mathematical transformation that operates on the time domain short segment interferograms in a frequency dependent manner. The result of the transformation suppresses frequencies not associated with the spectral features of interest. One type of digital filter, that is used on time domain interferogram segments [3], is a non-recursive FIR filter. The form of the filter function is given in equation (4).

$$Y(i) \; = \sum_{i=1}^{N} \; b(i) \; X(i\text{-}n) \tag{4}$$

Y(i) is the filter data point resulting from the summation of preceding X(i-n) raw data points that are weighted by N values of the b(i) filter coefficients. This filter strategy develops very narrow-bandpass filters for operation on the short interferogram segments. The digital filtering minimizes background effects leaving primarily the fundamental absorption and emission features of the target vapor cloud.

Another, type of digital filter is the recursive IIR filter. This filter in addition to the FIR term has a component that provides a feedback response. The functional form of the filter is given in equation (5).

$$Y(i) = \sum_{r=1}^{M} A(r) \; Y(r\text{-}n) \; + \sum_{i=1}^{N} B(i) \; X(i\text{-}n) \tag{5}$$

The first term in equation (5) is the feedback component because the

coefficients operate on previously filtered interferogram points.

IIR digital filters have one advantage and one disadvantage compared to FIR filters. The disadvantage is the instability when operating on an interferogram segment near the centerburst. The filter has an advantage for segments with lower dynamic range in that the feedback response generates very narrow-bandpass filters with few coefficients. Figure 3 shows the attenuation response of a 100 coefficient FIR filter versus a 12th order IIR filter. The number of computations for these two filter implementations is comparable. In both cases a very narrow-band filter is desirable for extracting the spectral feature centered at $939cm^{-1}$ associated with the simulant sulfur hexafluoride (SF_6). The plot in Figure 3 describes the filter frequency response, a logarithmic function (i.e. $-20log(Input/Output)$). The 12th order IIR filter shows a strong attenuation of -132dB for frequencies outside the region of interest. In contrast, the FIR filter shows a less significant attenuation of about -35dB with a much broader bandpass. The IIR filter for the SF_6 frequency in the short 150 point segment, displaced 150 points off the interferogram centerburst, gives greater attenuation of unwanted background and noise frequencies.

The result of filtering a short interferogram segment with the IIR filter is shown in Figure 4. Figure 4a shows the unfiltered segments containing SF_6 and air, while Figure 4b shows the corresponding filtered response with the narrow-bandpass IIR filter centered at $939cm^{-1}$. The presence of the fundamental frequency for SF_6 is apparent in the lower trace of Figure 4b.

A comparison of the two digital filters is shown in Figure 5. Figure 5 gives the magnitude response of sequentially collected filtered interferogram segments. An interferometer mounted on a U.S. Army UH-1 helicopter traveling at 120 knots and an altitude of 600 feet provided the set of sequentially collected interferograms. The downward looking interferometer flew over the cloud once in this data set. The boundaries of the SF_6 target cloud are defined by the short segment filtered response for interferograms 50 and 77. The magnitude response of the FIR and IIR filters are plotted. An improved signal-to-noise (SNR) response of approximately 25 percent for the IIR over the FIR is observed due to the individual filter frequency response characteristics.

Automatic detection of filtered interferogram segments for military scenarios requires pattern recognition. A linear discriminant function gives a separation of the filtered data into two classes: (1) data containing target cloud spectral features and (2) data without the target cloud spectral features. Linear discriminant functions are described in general by equation (5).

$$G(x) = Wo + \sum_{i=1}^{N} Wi\ Xi \qquad (5)$$

Wi is the weight vector; Xi is the response vector; Wo is the threshold vector; and G(x) is the discriminant function.

(A)

(B)

Figure 4.
The result of a digital filter for a short 150 point interferogram segment with an IIR filter is shown. The upper two traces are a plot of raw data and the lower two traces are filtered results.
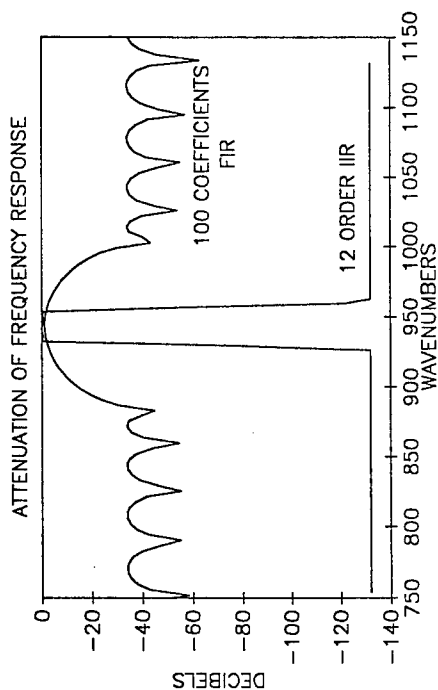

Figure 3.
Comparison of FIR and IIR filter frequency response necessary to extract the SF_6 feature at $939cm^{-1}$ using a comparable number of computations.
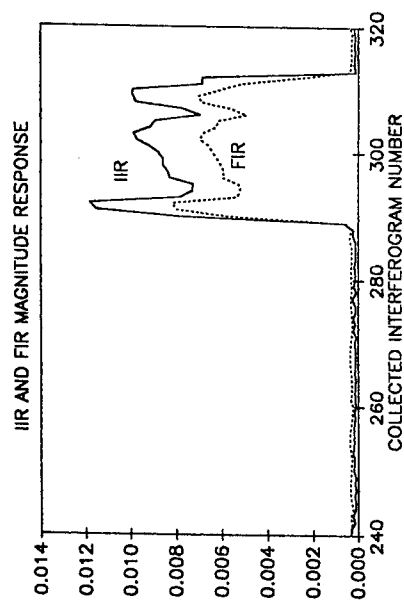

Figure 5.
FIR and IIR filter magnitude response for a single pass over a SF_6 vapor cloud with a helicopter mounted passive IR sensor.

The discriminant function defines an N dimensional plane (i.e. hyperplane) which intersects the data sample space. A convergence technique known as the perceptron convergence theorem permits calculation of the weight vector, Wi. This procedure calculates a weight vector for any misclassifications on the positive side of the hyperplane. The procedure is applied in a recursive manner until all samples are classified on the correct side of the hyperplane surface. A proper set of weight values minimize the response to unwanted spectral features while maximizing the response to the vapor target cloud features.

INTERFEROGRAM SEGMENT SELECTION:
    Interferogram segment location and size determine the instrument response (i.e. SNR) of a SSI. The ratio in the frequency domain of the spectral widths of the target vapor cloud species governs the interferogram segment selection. In a full scan interferometer, as the mirror moves away from the centerburst, a damping of the broad spectral features results. Broad-band spectral features damp rapidly while narrow-band frequencies damp slowly with mirror displacement. A maximum in intensity of the target spectral signature occurs as the background features diminish. Therefore, the ratio of the target cloud signature response to magnitude for the background radiation features must be maximized for an optimum interferogram segment selection.
    Interferogram data taken from a moving truck with a four wavenumber Michelson interferometer is shown in Figure 6. A segment of 150 points is selected at various positions out of 1024 point interferograms collected at 10KHz. This permits determination of the optimum interferogram segment location for compound signature identification. The segments are displaced away from the centerburst by 30, 150 and 240 points. These are labeled A, B and C respectively in Figure 6.
    Examination of each interferogram region shows that the number of ADC bits (i.e. resolution) is significantly less than that of the entire interferogram. The entire interferogram requires sixteen bit ADC resolution, while for segments A, B and C of Figure 6 the resolution is ten, nine and eight ADC bits respectively. Digitization of only the short interferogram segment permits a reduction of the dynamic range required of the ADC. The lower resolution ADC provides a higher accuracy in digitization of the short interferogram segment than the higher resolution ADC where only the least significant bits are used.
    The selected A, B and C short interferogram segments were digitally filtered with a thirty coefficient FIR filter for a set of about eight hundred interferograms. The stability of the FIR filter made it the filter of choice in determining the optimum segment location. The eight hundred interferograms were obtained with a ground mobile interferometer moving past a remote SF_6 source several times. The magnitude response of the filtered short segments provided an indication of the presence or absence of SF_6. The presence or absence of SF_6 as a function of interferogram number was visible

Figure 7.
Magnitude response of a FIR filter for three segment locations using approximately 800 interferograms is shown. During data collection the truck mounted sensor views a SF_6 vapor cloud several times.
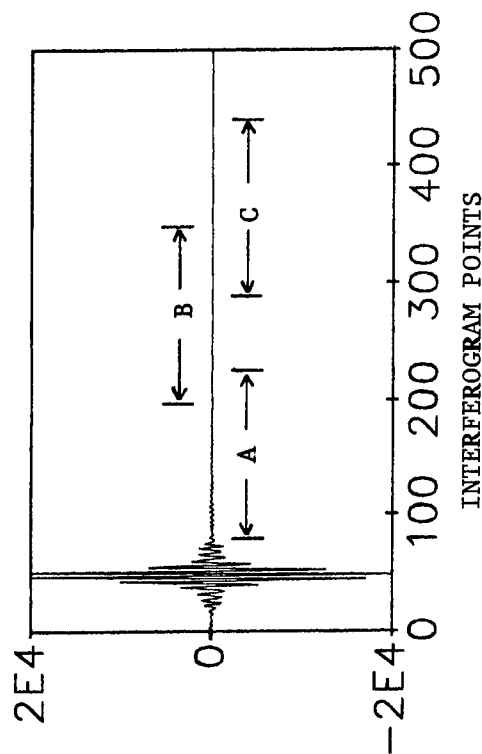

Figure 6.
Short 150 point interferogram segments are displaced by (A) 30, (B) 150 and (C) 240 points off the interferogram centerburst.

in the plots of magnitude response for the truck data in Figure 7. A measure of filter effectiveness (i.e. SNR) was reflected in the ratio of magnitude response for SF_6 present to absent. The SNRs for segment A, B and C were approximately 3, 50 and 20 respectively. Segment B showed the largest SNR, which was nearly seventeen times larger than segment A. The SNR of segment B was more than two times better than segment C. Therefore, segment B was the optimum of the three segments for SF_6 detection for this particular FIR digital filter.

INTERFEROMETER MIRROR VELOCITY:

In a SSI, mirror velocity is an important consideration for mobile applications. High mirror velocities tend to reduce interferogram smearing and give rise to a faster instrument response. Faster mirror velocities provide greater insensitivity to external vehicle vibrations. Determination of interferometer mirror velocity effect on toxic vapor detection requires consideration of the sensor response function listed in Table I. Table I records radiance measurements with two Michelson interferometers operating at comparable mirror velocities of 1.25 cm/sec and 1.58 cm/sec in columns 2 and 3 respectively. The Noise Equivalent Spectral Radiance (NESR) measures interferometer performance as an instrument response function.[4] The ratio that defines NESR is the theoretically known input blackbody radiation difference at two temperatures ratioed against the interferometer SNR output difference at the same two temperatures. Equation (6) gives the procedure for computation of the root mean squared noise component of the output SNR to be used in calculation of NESR.

$$N_{rms} = \frac{1}{\sqrt{2}} \sqrt{ \frac{1}{\bar{\nu}_2 - \bar{\nu}_1} \int_{\bar{\nu}_1}^{\bar{\nu}_2} [Spectrum_2(\bar{\nu}) - Spectrum_1(\bar{\nu})]d\bar{\nu} } \qquad (6)$$

$Spectrum_2(\bar{\nu})$ and $Spectrum_1(\bar{\nu})$ represent the FFTs of sequential interferometer scans at a given temperature. The spectral wavenumbers $\bar{\nu}_2$ and $\bar{\nu}_1$ describe the energy bandwidth used in measurement of the SNR. Equation (7) permits calculation of NESR at a particular wavelength.

$$NESR_{\bar{\nu}} = \frac{U_{\bar{\nu}}(T_2) - U_{\bar{\nu}}(T_1)}{SNR(T_2) - SNR(T_1)} \qquad (7)$$
$$\phantom{NESR_{\bar{\nu}} = \frac{U_{\bar{\nu}}(T_2) - U_{\bar{\nu}}(T_1)}{}}_{\bar{\nu} \qquad \bar{\nu}}$$

$U_{\bar{\nu}}(T)$ denotes the Planck equation that describes the radiance from a blackbody at temperature T. Large output SNRs for an input radiance level result in small NESR values on the order of 10^{-8}. The NESR values in Table I appear nearly identical for the two interferometers at all energies considered except $714 cm^{-1}$. The differences in detector response envelopes accounts for the discrepancy at $714 cm^{-1}$. The response envelope depends primarily on composition of the Hg:Cd:Te liquid nitrogen cooled detector.

372

TABLE I.NOISE EQUIVALENT SPECTRAL RADIANCE (NESR)

Wavenumber(cm-1)	NESR(Watts/cm2 sr cm-1)	
714	2.3	40
1000	1.5	1.6
1428	6.2	5.7

Note: All values are scaled by 10^{-8}.

The effect of mirror velocity on the detection of toxic vapors with a Michelson interferometer is given by equation (8).

$$SNR = \frac{\text{Signal Power}}{\text{Noise Power}} = \frac{U_{\bar{\nu}}(T) \cdot \theta \cdot \xi \cdot \Delta\bar{\nu}}{NEP/\sqrt{t}} \qquad (8)$$

Input radiant energy, interferometer parameters and detector response all contribute to the SNR. Terms in the numerator of equation (8) other than the input radiant energy comprise the interferometer parameters. Theta indicates the radiation throughput of the interferometer and accounts for various optical losses. $\Delta\bar{\nu}$ designates the interferometer resolution (i.e. reciprocal to total mirror travel). Xi represents the duty cycle of data collection from the interferometer. The detector observation time, t, depends on the interferometer mirror velocity. The denominator of equation (8) contains the detector response for modulated radiant energy from the interferometer. Noise Equivalent Power (NEP) is a measure of detector response for a given modulated input radiation level. Interferometer operation at mirror velocities (cm/sec) of (a) 1.58, slow; (b) 3.17, medium; and (c) 6.20, fast are considered in Table II.

TABLE II. SNR as a function of mirror velocity.

Ratio of Mirror Velocities	Inverted Ratio of SNR	
	Predicted	Measured
3.9 Fast/Slow	2.0	1.3
1.7 Medium/Fast	1.3	1.0
2.3 Fast/Medium	1.5	1.2

Table II allows a comparison of SNRs for the simulant SF6 at the aforementioned mirror velocities. If NEP is independent of mirror modulation frequency, then it follows, that the square root of the mirror velocity ratio is equivalent to the reciprocal of the SNR for the associated mirror velocities. The reciprocal SNR ratios in column 2 of Table II predicted from equation (8) are not equivalent to the measured values in column 3 of Table II. Column 3 of Table II shows that for various mirror velocities the interferometer maintains a constant SNR within the experimental precision of the measurements. The constant SNR implies one of two possible explanations. First, other noise sources in the signal conditioning electronics may dominate the Hg:Cd:Te detector noise. Second, NEP for a detector noise

373

limited sensor is a function of modulation frequency and tends to decrease to some constant value at an upper modulation frequency.[5] Determination of which case holds in these measurements requires minimization of various noise sources to insure that the predominate noise source is detector noise. Error source measurements are currently under investigation to determine which mechanism permits the SNR to remain constant with increased modulation frequencies.

DATA SAMPLING RATE:

Digitization rate is found to be another important parameter in the design of a mobile IR sensor. Signal averaging provides a means of improved target compound detection for a laboratory environment. This method rapidly reaches a point of diminishing returns, since the SNR improves only by the square root of the number of scan coadditions. To use the traditional signal averaging technique of coadding sequential files with an IR spectrometer requires a stable IR radiation source over the averaging time. Therefore, traditional signal averaging for a passive remote sensor necessitates the use of interferometer scan speeds significantly faster than any variation in the background terrain radiation. In a remote stationary mode of operation, the background radiation usually remains stable for eight interferogram coadditions. This coaddition gives a SNR improvement by a square root of eight. In a moving vehicle scenario, the background radiation varies rapidly precluding signal averaging by the traditional time sequential method. However, collection of more interferogram points over the same interferometer scan period gives an enhancement in the SNR equivalent to the time sequential method without the stable IR radiation restriction. Separation of the highly sampled interferogram into component 1024 point interferograms permits signal averaging in the frequency domain.

Current remote FTIR spectrometers collect data on every fourth positive fringe of the HeNe reference laser (i.e. 1024 point interferogram). Faster data acquisition electronics coupled with a zero crossing detector circuit enables data collection for every positive laser fringe (i.e. 4096 point interferogram) as well as every laser zero crossing (i.e. 8192 point interferogram). Table III shows the comparison of sampling rates with the ratio of SNR of a 4096 point interferogram collected at 40KHz to a 1024 point interferogram collected at 10KHz.

TABLE III. Signal to noise ratio for 40KHz versus 10KHz sampling rate.

Run #	SNR Ratio
#1	2.1
#2	2.4
#3	1.8
#4	1.7
Average	2.0

A 4096 point composite interferogram contains four interweaved 1024 point interferograms. FFTs of the individual 1024 point interferograms and coaddition yields a SNR improvement of two over a single 1024 point interferogram. Similarly, sampling at 80KHz provides an enhancement in the SNR of about three in comparison to a 10KHz sampling rate. Thus, a faster sampling rate improves the SNR regardless of the interferometer that is used. This improvement is significant for passive remote sensing applications, since a gain in the SNR over the Background-Limited Infrared Photodetector (BLIP) conditions is possible.[6] BLIP conditions represent the upper performance limit for a single scan passive sensor monitoring terrain background radiation.

Time domain processing of the short interferogram segments also show SNR improvement with higher sampling rates. The higher sampling rates combined with digital filtering reduces the errors due to aliasing.[7] If aliased noise is digitized, then its removal becomes impossible with signal processing. High frequency sampling increases the Nyquist sampling frequency and minimizes the effect of high frequency noise masquerading as a low frequency signal component. Elimination of aliasing requires a digital filter with a proportionally larger number of coefficients. Though computationally more intense this digital filtering approach offers flexibility and versatility not possible with analog filter techniques. For instance, the digital filtering technique may permit elimination of phase errors induced by the Butterworth analog bandpass filter. Examination of only the optical phase errors may allow determination of interferometer error sources and their contribution to photometric inaccuracies.

The effect of faster data sampling rates is also seen in dynamic analysis of the short interferogram segments. Dynamic analysis of the filtered interferogram segments indicates the presence or absence of spectral features with a trajectory. A plot of the filtered segment intensity against the rate of intensity change generates a trajectory. The trajectory for the presence of the single SF_6 spectral feature is circular, while the absence of SF_6 with only residual noise present produces a smaller random walk trajectory. Figure 8 shows the trajectories for SF_6 present to absent for different data sampling rates. A sampling rate of 10KHz in Figure 8 (a) and (b) is compared to a faster sampling rate of 40KHz in Figure 8 (c) and (d). The added resolution in the circular trajectory for a faster sampling rate offers an enhanced discrimination capability. The marked improvement for the higher sampling rate agrees with the previous SNR enhancement that is found in the frequency domain analysis.
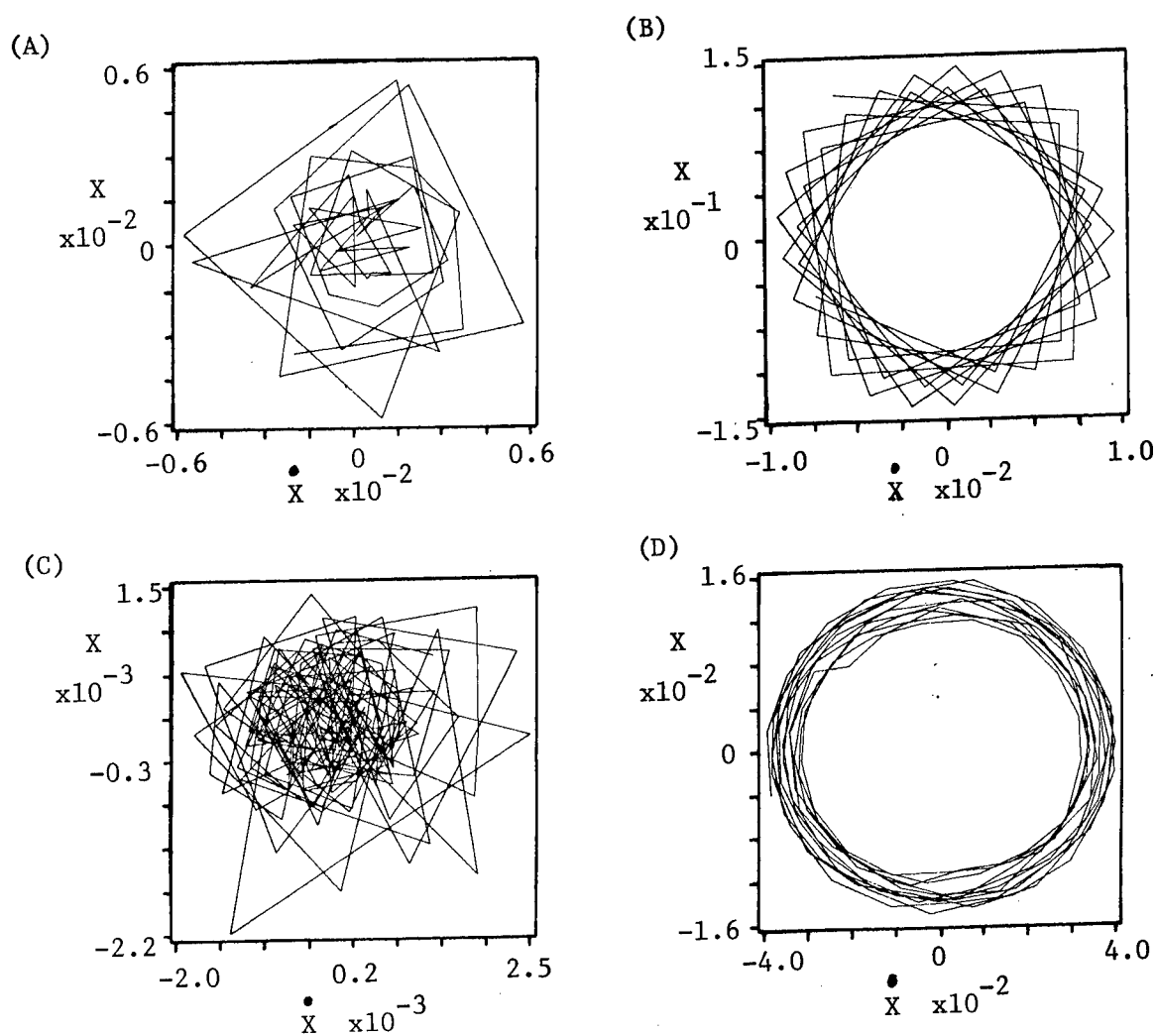
Figure 8. Dynamic analysis of data sampled at (A) 10KHz with SF_6 absent, (B) 10KHz with SF_6 present, (C) 40KHz with SF_6 absent, and (D) 40KHz with SF_6 present, is shown.

CONCLUSIONS:

The development of a short scan interferometer addresses the scenario of a highly mobile battlefield environment. The unique short-scan interferometer utilizes only a short interferogram segment in contrast to a full scan interferometer. Selection of the short intermediate interferogram segment is shown to reduce the noise and the effects of the rapidly changing background frequency contributions for mobile ground vehicle data. Subsequent digital filtering of the selected short segment retains only the spectral signature of the target vapor cloud. An optimum filter response for helicopter field data is shown by implementation of an IIR filter. Faster scan rates minimize smearing of spectral features. Increased mirror velocities also provide an inherent immunity to vehicle vibration. Higher data sampling rates enhance the detection of a toxic vapor cloud in a mobile environment with an approach not possible with traditional signal averaging techniques. These considerations, along with the critical digital signal processor for real-time instrument response, provide the basis of a novel passive IR mobile sensor.

REFERENCES:

[1] D.F. Flanigan and A. Walter, "The Computer Simulation and Optimization of Passive LOPAIR", ED-TR-74070, JAN 1975.

[2] R.T. Kroutil, J.T. Ditillo and G.W. Small, "Signal Conditioning Techniques for Remote Infrared Sensing" in Computer Enhanced Analytical Spectroscopy (H. Meuzelaar, ed.), Plenum Press, NY, Chapter 4 (1990).

[3] G.W. Small, R.T. Kroutil, J.T. Ditillo and W.R. Loerop, "Detection of Atmospheric Pollutants by Direct Analysis of Passive Infrared Interferograms", Analytical Chemistry, vol. 60, 264 (1988).

[4] B.J. Vastag and S.R. Horman, "Calibration of a Michelson Interferometer Spectrometer", Fourier Transform IR Spectroscopy Proc. SPIE, vol. 289, 74 (1981).

[5] P.R. Griffiths, J.A. deHaseth and L.V. Azarrage, "Capillary GC/FTIR",Analytical Chemistry, vol. 55, 1361A (1983).

[6] R.H. Kingston, "Detection of Optical IR Radiation", Springer-Verlag, Berlin, 20 (1978).

[7] L.J. O'Connell, "Apparent Effects of Aliasing in Time Domain Processing", Proc. Digital Signal Processing Symposium, Sandia Laboratories Sand77-1845C, Albuquerque, NM, p.160, DEC 1977.

A General Synthesis of Side Chain Liquid Crystalline Polymers (U)

John H. Cornell
Soldier Science Directorate
U.S. Army Natick RD&E Center
Natick, MA 01760-5020

Charles A. Lovelette
Chemistry Department
St. Anselm College
Manchester, NH 03102

INTRODUCTION

The objective of this project is to develop eye protection media which will defend the soldier against low energy laser threats such as laser range finders and target designators. Recently a new battlefield threat has emerged through the potential use of tunable laser weapons systems. Coherent radiation from these sources even at low energy levels is capable of inflicting severe damage to the retina of the eye. One mode of protecting against such threats involves the development of nonlinear optical polymers, which permit the transmission of visible light at ambient intensities but will reversibly block visible and near infrared radiation at intensities harmful to the eye.

There has been great interest recently in nonlinear optical materials because of the growing need in the telecommunication and computer industries for high speed, wide-bandwidth optical switching and processing devices. Nonlinear optics involves the interaction of incident electromagnetic radiation with various media to produce emergent radiation altered in phase, frequency, amplitude or other characteristics from the incident radiation. In principle it should be possible to design devices that make use of one or more of those changes in the transmission characteristics of the nonlinear optical media to secure the desired eye protection. To attain this objective, optical materials that give nonlinear effects, such as a change in refractive index of the media, are required. Successful candidate materials must produce such effects with sufficient magnitude and speed to block the incident laser pulse. Devices comprising nonlinear optical components for individual eye protection should result in reduction of pulse intensity equivalent to an optical density of four within one nanosecond.

In recent years organic and polymeric materials have emerged as promising classes of nonlinear optical media because in these materials the nonlinear polarization originates from purely electronic effects.[1] This leads to ultrafast (subpicosecond) response, broadband operation, and large nonresonant optical nonlinearities due to electron delocalization. In addition these materials possess excellent chemical and structural stability and do not require environmental

protection or cryogenic operation. Polymers of pyrrole, thiophene, benzene, aniline, benzoxazole, and benzothiazole can be fabricated as self-supporting films, filters, coatings, monolayers, liquid crystalline polymers, etc. [2-6] They are susceptible to structural variation by chemical synthesis or processing and possess a high potential for larger scale production through existing polymer manufacturing techniques.

At Natick RD&E Center a program for the development of nonlinear optical polymers that can be incorporated into devices for laser eye protection is in progress. The nonlinear polymer should pass visible light at ambient intensities but reversibly block laser radiation at intensities harmful to the eye. The program for the development of nonlinear optical polymer for use in eye protective devices is divided into two phases (a) selection and synthesis of polymers that potentially have suitable nonlinear optical properties and (b) evaluation of the nonlinear optical properties of candidate polymers. A facility for the measurement of $\chi^{(3)}$ is being constructed at Natick, Research Development and Engineering Center. A neodymium dye laser will be used and $\chi^{(3)}$ determinations carried out using the degenerate four wave mixing technique. When feedback is available from this facility, it will of course be used to explore the relationship between $\chi^{(3)}$ activity and polymer structure. This data in turn will be employed for the design and synthesis of polymers that will display the highest third order nonlinear optical effects.

Current out-of-house efforts center on polybithiophenebenzylidene and substituted polybithiopheneacetoxybenzylidene [7].

Prior to discussing the selection of candidates for nonlinear optical polymers for laser eye protection, the principles involved in providing the type of protection will be outlined.

BACKGROUND

The minimum performance requirements for laser eye protection are a pulse intensity reduction of at least 10^4 times with a response time of 1 nanosecond or less. Normal chemical reactions involving the movement of atoms cannot be expected to yield a sufficiently rapid response. On the other hand response times as short as 10^{-15} seconds are possible where only the movement of electrons is involved. Large classes of conjugated molecules and polymer structures exist whose nonlinear optical responses of this type occur by lossless excitation of the pi-electron states, especially those containing large charge correlations.

When an electromagnetic impulse is propagated through a medium, the polarization (P) of the medium is expressed in terms of the field strength E:

$$P = \chi^{(1)} E + \chi^{(2)} E^2 + \chi^{(3)} E^3 + \ldots\ldots$$

The first order term ($\chi^{(1)}$) is associated with linear transmission at ordinary light intensities. As light intensities increase, nonlinear effects associated with the second order ($\chi^{(2)}$) and third order ($\chi^{(3)}$) terms become more significant.

The higher order nonlinear effects are associated with certain symmetry requirements. Thus $\chi^{(2)}$ is observed only with non-centrosymmetric systems while $\chi^{(3)}$ is characteristic of centrosymmetry. The higher orders also produce different physical effects on incident radiation. $\chi^{(2)}$ results in frequency doubling (second harmonic generation) and $\chi^{(3)}$ in frequency tripling and changes in the refractive index. It is this last property that is considered most significant from the standpoint of individual laser eye protection.

An initial selection of candidate nonlinear optical polymers was made on the basis of the molecular structures of compounds that have exhibited nonlinear optical properties. These structures include extended pi-electron systems, permanent dipoles, and anisotropic structures.

Moieties containing these structural features may be distributed in various ways throughout the polymer chain, giving rise to different types of nonlinear optical polymers. These optical polymers comprise polymers in which the main chain is fully conjugated, or in which the centers of unsaturation are still in the main polymer chain but are broken up into segments by saturated moieties interspersed along the chain. Finally, the centers of unsaturation may be in the form of pendent groups suspended from the main chain by saturated spacer groups.

RESULTS AND DISCUSSION

The synthetic methods developed in this program will facilitate the preparation of side chain liquid crystal polymers with the following structural variations: polymer, backbone, restricted and unrestricted rotation at the site of side chain attachment to the backbone, length and flexibility of the spacer segment, and the type of mesogen incorporated. Systematic variation of the polymer structure in this manner will permit a detailed evaluation of the effect of such variation on the nonlinear optical properties for that type of polymer.

One method, Scheme 1, starts with acrylic acid and one of its homologs and/or derivatives, e.g., methyl acrylate or methyl methacrylate. A bifunctional spacer group comprising methylene and aryl units terminating with amino and/or hydroxy groups is attached by esterification or amination of the acrylate carboxyl function. A mesogenic group, e.g., cyanophenol, is then fixed to the terminal end of the spacer. Finally the modified acrylate monomer is polymerized using conventional catalyst systems to give a side chain polymer having the structural variation that is required.

Starting with readily available monomeric vinyl acids or methyl esters (1) represented an attractive beginning and formed the basis of the entire synthetic approach. Initial experiments centered on an attempt to accomplish ester interchange reactions with the appropriate methyl esters and terminal diols using type A molecular sieves [8]. This procedure appeared quite attractive in the context of experimental simplicity allowing the conversion to be accomplished without any special equipment. However, the procedure was unsuccessful, affording only clear, highly viscous liquid products whose infrared spectra contained OH and aliphatic CH stretching as well as C=O stretching but lacked the all important C=C stretching frequency. Proton nuclear magnetic

resonance confirmed the absence of the double bond, and failed to exhibit any vinyl proton resonances. Further attempts to characterize these materials were unsuccessful and the method was abandoned.

SCHEME 1

$$CH_2=\overset{\overset{\displaystyle R}{|}}{C}COOH \quad + \quad HO-(CH_2)_x-OC_6H_4COOH$$

$$(1) \qquad\qquad\qquad (2)$$

TsOH,
hydroquinone

$$CH_2=\overset{\overset{\displaystyle R}{|}}{C}COO-(CH_2)_x-OC_6H_4COOH \qquad (3)$$

1.) thionyl chloride, DMF
2.) Mesogenic anion (M)

$$CH_2=\overset{\overset{\displaystyle R}{|}}{C}COO-(CH_2)_x-OC_6H_4COOM \qquad (4)$$

polymerization

$$\overset{\displaystyle +\!\!\left(CH_2-\overset{\overset{\displaystyle R}{|}}{\underset{|}{C}}\right)_{\!n}}{COO-(CH_2)_x-OC_6H_4COOM} \qquad (5)$$
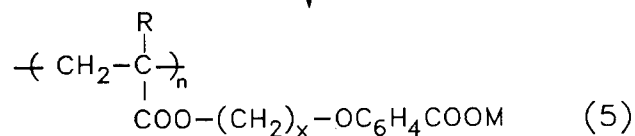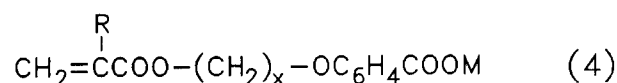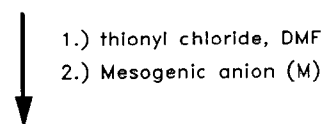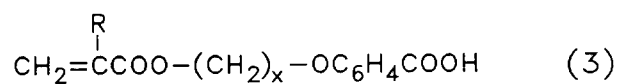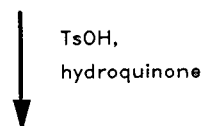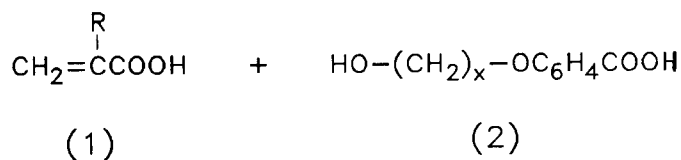
Next, standard Fischer esterification procedures were employed using the protocol described by Ringsdorf [9]. This method using the corresponding acids was successful, affording a series of esters whose structures are illustrated by (3): R=H,CH$_3$ and X=2,3,6. The presence of the aryl carboxy group served two purposes: its acidic nature made purification simple and it served as a site for attachment of the mesogenic group. This was most readily accomplished by conversion into the acid chloride with thionyl chloride, followed immediately by the addition of the appropriate nucleophile, e.g., p-cyanophenol. As is readily apparent, mesogenic group variability is, in theory, very extensive using this approach.

Finally this general approach affords a convenient pathway for the introduction of an amide-type functional group linking the main-chain backbone and the spacer functionality. This linkage serves the purpose of introduction of restricted rotation about the C(=O)NH bond. This is accomplished from the commercially available acid chlorides (1) where R=H,CH$_3$. Treatment of the acid chlorides with hydroxyamines gives the desired amide bond. The hydroxy group is then converted to the tosyl ester and treated with the dianion of p-hydroxybenzoic acid, allowing the preparation of a completely analogous series of monomers with the property of spacer rigidity.

Further rigidity is possible by treating (1) with diamines of varying length. Also, one could employ a cyclohexane ring spacer by utilizing 1, 4-cyclohexanediol. The basic goal of establishing a simple method for the synthesis of side chain liquid crystals having a wide variety of controlled variables was accomplished. This establishes the framework for an extensive program in this area with the goal of discovery of nonlinear optical activity in such systems.

Another synthetic route that involves the chemical modification of preformed acrylate polymers has been developed, Scheme 2. In this case polymethyl methacrylate is treated with a bifunctional amine such as ethanolamine or ethylenediamine, resulting in the displacement of the methoxy group and the formation of an amide linkage. Attachment of a mesogenic group to the reactive terminal end of the spacer group will give the desired side chain nonlinear candidate polymers. Polymers obtained in this way will be compared to those prepared according to Scheme 1, which have the same structure.

SCHEME 2

$$-\!\!\left(CH_2-\underset{\underset{COOCH_3}{|}}{\overset{\overset{CH_3}{|}}{C}}\right)_{\!n} \qquad (1)$$

ethylenediamine   aminoethanol

$$-\!\!\left(CH_2-\underset{\underset{CONHCH_2CH_2NH_2}{|}}{\overset{\overset{CH_3}{|}}{C}}\right)_{\!n} \quad (4)$$

$$-\!\!\left(CH_2-\underset{\underset{CONHCH_2CH_2OH}{|}}{\overset{\overset{CH_3}{|}}{C}}\right)_{\!n} \quad (2)$$

TsOM   TsOM

$$-\!\!\left(CH_2-\underset{\underset{CONHCH_2CH_2NHM}{|}}{\overset{\overset{CH_3}{|}}{C}}\right)_{\!n} \quad (5)$$

$$-\!\!\left(CH_2-\underset{\underset{CONHCH_2CH_2OM}{|}}{\overset{\overset{CH_3}{|}}{C}}\right)_{\!n} \quad (3)$$
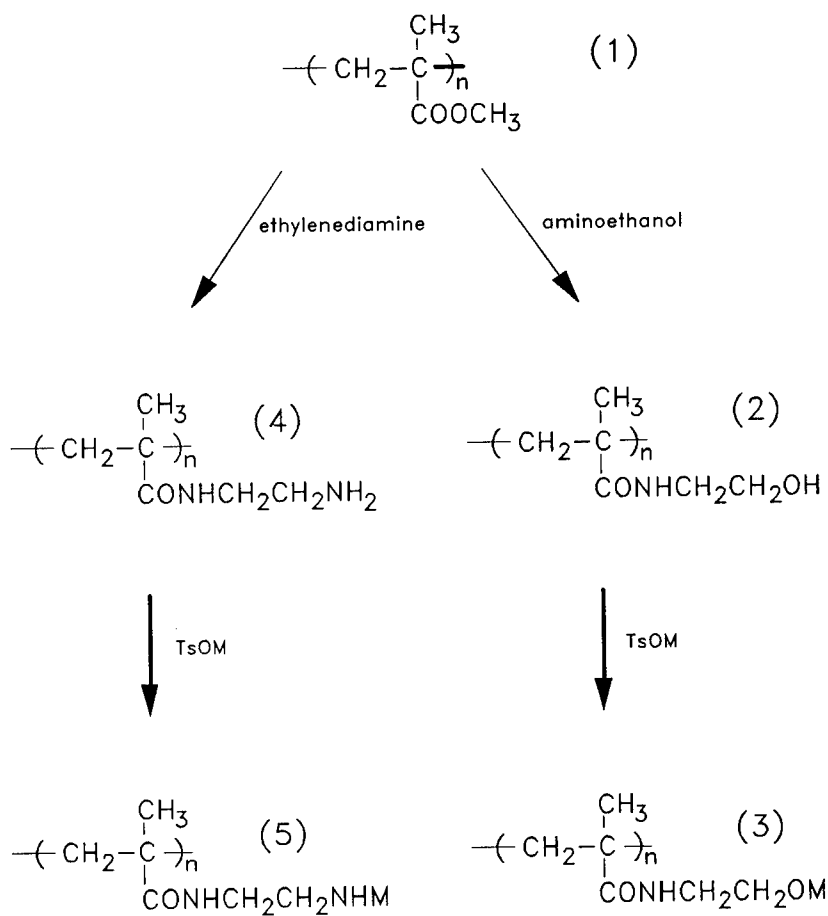
EXPERIMENTAL

Preparation of Compounds (2): (Scheme 1)

In a typical experiment, 4-hydroxbenzoic acid (13.8g, 0.1 mole) was dissolved in ethanol (100mL) and potassium hydroxide (15g, 0.3 mole) was added along with a catalytic amount of potassium iodide. The solution was heated and stirred while the chloro alcohol (0.1 mole) was added dropwise. The mixture was refluxed for 15 hours, then cooled to room temperature and the solvent removed under reduced pressure. The semisolid residue was dissolved in water (250mL) and the aqueous solution made acidic with hydrochloric acid (6M). The precipitate was collected and recrystallized from ethanol. (See Table 1.)

Preparation of Compounds (3): (Scheme 1)

A mixture of the 4-hydroxybenzoic acid (0.25 mole), acrylic, or methacrylic acid (1 mole), dichloromethane (200mL), p-toluenesulfonic acid (1g) and hydroquinonone (10g) was refluxed overnight in a Dean-Stark apparatus. If the theoretical amount of water had not been collected the reflux was continued until sufficient water was collected. The cooled reaction mixture was diluted with diethyl ether (1500mL), washed with warm water and dried over magnesium sulfate. Removal of the solvent followed by recrystallization from isopropanol afforded the final product.

Preparation of Compounds (4): (Scheme 1)

Compounds 3a-c (0.03 mole), dimethylformamide (1mL), 2,6-di-t-butyl-4-methylphenol (0.5mL) and thionyl chloride (20 mL) were stirred for 30 minutes at room temperature. After the reaction period the thionyl chloride was removed, first at the water aspirator and finally under high vacuum (1-1.5h). The residue was dissolved in dry ether (50mL) and added to a solution of the mesogen (0.03 mole) and triethylamine (5mL) in dry tetrahydrofuran (100mL) at 0°C. After warming to room temperature, the solvent was removed and the solid residue was dissolved in dichloromethane, washed with water and dried over magnesium sulfate. Final purification was effected by silica gel column chromatography using dichloromethane as the mobile phase. Removal of the solvent afforded 4a-c (Scheme 1).

Preparation of Compound (2): (Scheme 2)

PMMA (med. MW, Aldrich 18,224-9), 3.0g (0.03 mole) was refluxed and stirred for 21 hours with 10mL 95% ethanolamine (0.16 mole). The resulting thick liquid was dissolved in methanol and transferred to a distilling apparatus fitted with a Dean and Stark trap. Excess toluene was added and the mixture stirred and refluxed until the formation of a second layer in the Dean and Stark trap ceased (ca. 20 h). The toluene was decanted from the pot and the glassy residue washed several times with toluene.

The product was dissolved in refluxing methanol and was clarified by centrifugation. Excess tetrahydrofuran (THF) was added to the stirred supernatant, which resulted in the precipitation of a white solid. The product was decanted and stirred with THF several times and collected by vacuum filtration. It reached constant weight after drying in air for two weeks, 2.5g (64.6% yield).

The appearance of strong IR absorption at 1625 cm^{-1} (amide CO) and 3320 cm^{-1} (OH) and the weakening of the 1725 cm^{-1} band (ester CO) in the product indicated that the reaction had proceeded as shown in Scheme 2.

The foregoing experiment was repeated with low MW PMMA (Aldrich 18,223-0) to yield (2) (80.1 % yield). When this preparation was repeated with high MW PMMA (Aldrich 18,226-5), the immediate product was a solid insoluble in refluxing methanol, dimethylformamide and tetrachloroethylene.

Preparation of Compound (4): (Scheme 2)

PMMA (med. MW) was treated with ethylenediamine using the same procedure as for ethanolamine. The product (2) was obtained as two fractions: (2a) soluble in refluxing methanol (33.8% yield), and (2b) soluble in refluxing tetrachloroethylene (15.6% yield).

TABLE 1. Synthesis of Acrylate Monomers

Compound Scheme 1	Substituent	Monomer Chain Length	Mesogenic Group	Percent Yield	IR Bands (KBr, cm^{-1})
2a		x=2		51	3350, 1680, 1610, 1440, 1255, 775, 650
b		x=3		57	3380, 3290, 2980, 2890, 1680, 1610, 780
c		x=6		73	3400, 2940, 1690, 1610, 1360, 780, 645
3a	R=H	x=2		52	3010, 1725, 1690, 1610, 1410, 1360, 1270
		x=3		51	
		x=6		60	3010, 2940, 1720, 1690, 1610, 1360, 1270
3b	R=CH$_3$	x=2			3320, 1685, 780, 645
		x=3			3410, 1690, 780, 650
		x=6			3350, 1700, 790, 670
4a	R=H	x=2	M$_1$	35	2230
4b		x=3	M$_1$	42	2230
4c		x=6	M$_1$	40	2230
4d		x=2	M$_2$	45	1690
4e		x=3	M$_2$	33	1690
4f		x=6	M$_2$	38	1690

$$M_1 = p\text{-}HOC_6H_4CN$$
$$M_2 = p\text{-}HOC_6H_4COOH$$

SUMMARY AND CONCLUSION

A simple synthetic method has been created permitting the preparation of side chain liquid crystalline polymers with the following structural variations: polymer, backbone, restricted or unrestricted rotation at the site of the side chain attachment to the backbone, the length and flexibility of the spacer segment, and the type of mesogen incorporated. This method permits a detailed evaluation of chemical structure-liquid crystal properties for this type of polymeric system.

The synthetic method starts with acrylic acid or one of its homologs and/or derivatives, e.g., methyl acrylate or methyl methacrylate. A bifunctional spacer group comprising methylene and aryl units terminating with amino and/or hydroxy groups is attached by esterification or amination of the acrylate carboxyl group. A mesogenic group, e.g., p-cyanophenol, is then fixed to the terminal end of the spacer. Finally the modified acrylate monomer is polymerized using conventional catalyst systems to give a side chain polymer having the structural variation that is needed.

Another synthetic route that involves chemical modification of preformed acrylate polymers has been developed. Polymers obtained in this way will be compared with those prepared by the first route having the same molecular structure.

REFERENCES

1. William, D.J., Angew. Chemie, Intl. Ed. 23, 690-703. (1984)

2. Sato, M., Tanaka S. and Kaeriyama K., J. Chem. Soc., Chem. Commun. 1985, 713-714.

3. Bjorklund, R. and Lindberg, B., J. Chem. Soc., Chem. Commun. 1986 1293-1295.

4. Bianco, P., and Musiani, M., J. Appl. Poly. Sci. 26, 4247-4257 (1981).

5. Tourillon, G. and Garnier, F., J. Electroanal, Chem. 135, 173-178 (1982).

6. Feinberg R., Optical Engineering Report, Soc. Photooptical Instrumentation Engineers, July 1987, 5A-7A.

7. Contract DAAK60-84-C-0044, Honeywell Systems Res. Center, Minneapolis, Minnesota.

8. Roelofsen, D.P., Hagendoorn, J.A., van Bekkum, H., Chemistry and Industry, 1622 (1966).

9. Portugall, M., Ringsdorf, H., Zentel, R. Makromol. Chem., 183, 2311 (1982)

Program Support Environment
for the Theater Construction Management System (U)

*Kenneth H. Crawford, Mr.
Kendra Z. Azzi, Ms.
Biju Kalathil, Mr.
U.S. Army Construction Engineering Research Laboratory
P.O. Box 4005
Champaign, Illinois 61824-4005

Introduction

The engineers' role in the Theater of Operations (TO) is both unique and multifaceted. In the Theater of Operations, engineers are charged with the construction and repair tasks associated with functions performed in the theater. These include establishing and maintaining lines of communications (LOC), troop beddown, hospital and airfield construction, and logistic facilities. Various aspects of the theater construction process have been automated, but there is no integrated theater construction management system available for the Army engineer. Integration as used here means that a functional area within the construction process will be able to use data generated by other functional areas. The achievement of this goal requires the existence of computer software that has the capability of managing large collections of data residing in various databases which may use incompatible methods for data storage. To fill this void and to achieve this integration, the U.S. Army Construction Engineering Research Laboratory (USACERL) is developing the Theater Construction Management System (TCMS) and the Program Support Environment (PSE). TCMS will be a large collection of construction management and engineering calculation software packages integrated by data links that permit data sharing between functional areas of the construction management process. Figure 1 illustrates the complexity of data integration required for TCMS. Figure 1 should not be interpreted as a final model for TCMS; it was developed during the research phase of TCMS development. The final TCMS model may be quite different. This figure does, however, illustrate the massive amount of data sharing required between functional areas of construction management.

PSE is the vehicle which will permit the data integration within TCMS. While researching the problem of data management for TCMS, the team at

Figure 1. Theater Construction Management System – Data Flow Model
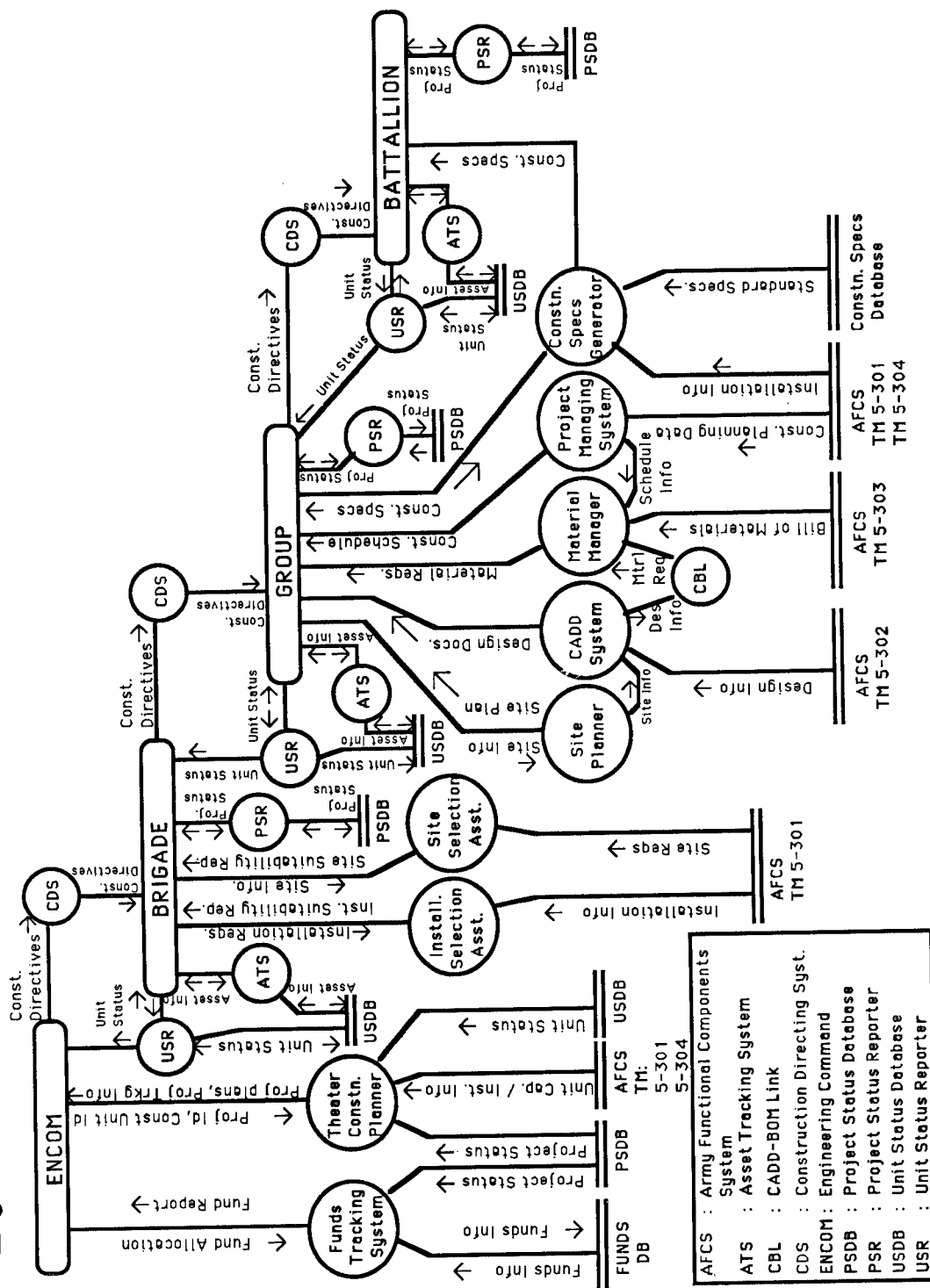
USACERL responsible for the development of TCMS concluded that software did not exist that possessed the required capabilities to fulfill the demands of TCMS. For example, the reporting functions performed by the Army engineer require that data from several databases be included in reports. As simple as this application may seem, standard database management packages do not provide this capability. A second example is the need to extract data from one database, say the resources available data, and convert it to a format readable by a project scheduling software package. There are numerous additional examples that could be cited to justify the development of PSE. It was clear to USACERL researchers that a database software package had to be developed which could not only meet the known demands of TCMS data management and integration, but which could be readily expanded to meet future demands. This conclusion led to the birth of PSE. Although PSE contains all of the capabilities of a standard database management package that can be purchased off-the-shelf, it is also much more. It is a library of software which gives systems developers powerful tools for expanding database management concepts and integrating software packages that are purchased off-the-shelf.

The first major decision in the development of PSE was the determination of which data storage format should be used for actual disk storage of data. This was an important decision because the existence of incompatible storage formats is the major cause of many of the incompatibility problems between commercial computer systems. In order to minimize these incompatabilities, it was decided to use the most widely utilized data storage format in commercial off-the-shelf packages. This particular format, though perhaps not the most responsive ever conceived, yields very acceptable response times to queries, was easy to program, and minimizes the data conversion necessary to attach data to PSE that originates outside TCMS. To enhance response times, the user can create an index on any field in any database. PSE software will automatically create indices on selected data fields to increase efficiency on commands that require a significant amount of computer time and disk access time.

Current Status of PSE

PSE, as currently developed, includes the following functions and capabilities:

1. A database management system is available which permits the user to attach to PSE a database created in a commercial database package. Once attached, the database becomes an integrated part of PSE. If necessary, data conversion is performed. Data elements are placed in the PSE data dictionary where they are tracked. PSE contains a complete collection of the usual commands for creating, deleting, querying and updating databases. The capability of attaching an existing database to PSE has the advantage of

391

permitting existing data to be integrated without manual reentry. Once attached, the data can be manipulated using PSE data manipulation commands.

2. All data attached to PSE is automatically placed in an indexed data dictionary which permits the user to easily locate all databases which contain a particular data element. The data dictionary monitors changes to data to insure that all occurrences of data elements reflect updates. The data dictionary will contain a Hypertext query capability which will greatly enhance the users' ability to examine and track data elements.

3. A Structured Query Language (SQL) contains the standard SELECT, INTERSECT, JOIN, UNION, PROJECT and SUBTRACT commands found in relational database management systems. This is a very powerful capability. Different databases are considered to be relations by PSE and the SQL commands present the user with powerful commands to manipulate data. With these commands the user can form new databases utilizing data from existing databases, regardless of the original source of the data. For example, by projecting several fields from one database and additional fields from a second database, the user can join the two resulting databases on a common field to form a third database containing the selected fields from the original two databases. The original databases remain intact. Although several commercial database packages contain relational operators for manipulating tables of data, PSE is unique in its ability to perform these operations on complete databases.

4. An interactive ASCII file to database file converter permits the user to specify the location in an ASCII file where fields in a database schema reside. PSE then loads the ASCII file into the specified database. This facilitates conversion of data which originates from COTS software in ASCII file format, a common commercial method for transferring data between software packages.

5. A very powerful report designer facilitates the design of reports that are specific to a particular application. The report designer is interactive with menu and cursor. The user enters commands at selected locations on the report and these commands will execute when the report is generated. These commands are SQL commands which cause PSE to query the named database for the data specificied in the command. Each command can query any database which is attached to TCMS, thus permitting reports to contain data from an unlimited number of databases. There are commands for totaling data fields with or without listing the data. Additional commands permit accumulating totals. The report designer is particulary useful in the Army since it permits each unit to have its own report format which can be readily changed. The report designer also permits the user to accumulate totals as the reports progress up the chain of command.

6. A report library is available to store report designs. A report design stored in the library can be instantly recalled for updating, revising, or printing the report. Library entries can be transmitted to other machines for use.

7. The report generator will call a report design from the library upon user command and generate the specified report. The report is placed in an ASCII file for examination before being sent to the printer or transmitted electronically. The report generator contains a command interpreter which translates and executes commands in the report design. This interpreter calls upon the SQL library of query commands to execute the database queries in real time.

8. A software development library is available for future expansion of TCMS. As new functional areas are identified that should be addressed by TCMS, the software development team will first attempt to identify COTS software that adequately addresses the functional area. If this is unsuccessful, software will be developed. In either case, the software will need to be integrated with the existing TCMS family of software. The major tasks here will be the conversion of data from the new software for export into TCMS and the conversion of TCMS data for import into the new software. The PSE software development library contains routines to expedite this development. TCMS is being developed in a stepwise fashion, that is, a prototype version with limited capability will be placed in the field at selected test sites for three primary purposes. First, interaction with the user and the prototype will assist researchers in identifying those areas of Army construction management that could be best served by automation. Second, the prototype will assist in beginning the automation process in the field. This will help to identify those areas where training is required. Finally, the software in the prototype needs to be tested for errors and human interface design. The software development library will facilitate the growth and expansion of TCMS once it has been placed in the hands of the user. This library will contain a menuing and windowing capability. These will permit developers to tailor the human interfaces to individual units.

9. A password system is available for limiting access to databases attached to TCMS. These passwords can give several levels of protection from total lock-out to read-only of selected fields to complete access. The password protection for most databases would be set by the user. There will be databases attached to TCMS that will have read-only capability for all TCMS users. These databases are maintained by other Army agencies.

10. A data distribution capability will permit selected data to be transmitted to subordinate units. It is expected that the Engineering Command in the Theater of Operations will be the unit responsible for the totality of data contained in TCMS. Subordinate units will need only

selected portions of data which may reside in several databases. PSE will have the capability to selectively choose data for distribution to brigades. Brigades will be able to further break down the data for selective distribution to battalions, and likewise battalions to companies.

11. An extended tutorial capability is available to assist the user in utilizing the commands in the PSE. The final soldier machine interface for TCMS will not be completed until users in test units provide feedback which will assist developers in providing the best possible interfaces to assist soldiers in using TCMS. Due to the anticipated complexity of TCMS, it is the opinion of the developers that PSE should not add to this complexity, thus a developmental goal has been that PSE be usable by those with a minimum of computer training. To achieve this goal, a considerable amount of time has been spent in creating a help system for PSE. Each command in the PSE is accompanied by a help screen or a help command. Training will be necessary for users of TCMS, especially on the satellite subsystems such as computer aided drafting, project scheduling, and site planning.

Development Environment for PSE

PSE is being developed on personal computers running the Disk Operating System (DOS) with an EGA (Enhanced Graphics Adaptor) monitor or a VGA (Virtual Graphics Adaptor) monitor. This environment was chosen because this class of machine is most readily available in the field at present. A completely transportable high level language is being used in the development. The language chosen is highly compatible with the Unix operating system. Unix is the operating system of the Army Command and Control System (ACCS) common hardware. Transporting of PSE to this common hardware will require minimal effort because the DOS dependent functions are being isolated for easy replacement by corresponding Unix functions.

PSE and TCMS will satisfactorily operate on DOS class machines provided these machines meet minimum requirements. Requirements consist of a minimum of 640 kilobytes of random access memory (RAM) and a numeric coprocessor. The coprocessor is needed to assist with floating point operations in subsystems of TCMS such as computer aided drafting; PSE does not require the coprocessor. Since PSE is designed to handle an unlimited amount of data, and since TCMS will require large data storage, a higher capacity hard disk than the standard 20 or 30 million byte capacity is recommended. Development of PSE software has utilized the modular approach which means that the software consists of a collection of modules which perform non-overlapping functions. Modularity permits overlaying of modules in memory. This technique permits the sharing of RAM by several modules which are never required to be resident in memory at the same time. As a result, a very large computer system can successfully operate in 640 kilobytes of RAM, even though the total system size may exceed that amount of memory. The operating

system accomplishes this task by loading modules from the hard disk when needed. This procedure can be significantly enhanced if the modules reside in memory above the base 640 kilobytes of RAM and are loaded into base memory from there rather than from the hard disk. Data transfer is significantly slower from hard disk than from memory. Thus, computers with extended memory above 1 million bytes of RAM will execute PSE and TCMS more rapidly. Extended memory will be used automatically by PSE if it is available on the machine.

Future Development of PSE

The capabilities and functions of the PSE described above have been coded and tested in the laboratory, but field testing is necessary before a final version will be ready. It is a stand alone system that will serve as the foundation of TCMS. In fact, several packages have already been integrated with PSE in the effort to develop TCMS. These include computer aided design, Army Facility Component System (AFCS), project scheduling, and Theater Army Construction Planning System (TACAPS). PSE has performed satisfactorily in these experiments. PSE has potential uses in other areas of research and development. It has been used in two other applications at USACERL which required data management capabilities.

Immediate plans for PSE call for further testing and refinement of existing code. It is anticipated that when field operating test units begin to use the software, several desirable functions will be identified. These will be developed as they are identified.

Future plans call for PSE to be used as a basis for further development in the data management arena. Experiments are underway to add object storage capability to PSE. Object oriented programming is considered by many to be the programming methodology of the future. Objects created in one computer software system can be utilized in other software systems, thus expediting systems development and reducing soaring software development costs.

Another avenue of investigation is the addition of intelligent capabilities to PSE. For example, when an engineer in the field is attempting to select a site for an installation, an intelligent database could narrow the choices by eliminating sites that did not meet minimum criteria by examining appropriate data such as terrain data and installation requirements.

Summary

The Program Support Environment is a collection of computer software modules which will permit data originating from several distinct sources to be integrated into one massive collection of databases. The resulting

collection of data appears to the user as though it was one database. PSE provides the user with commands to manipulate this data to create new databases. It contains data integrator modules for transforming data storage formats to facilitate data transfer between software packages. PSE is a stand alone collection of software which could serve as the database component of any software development. It will be used in developing the Theater Construction Management System.

Conclusions

The software for the PSE was developed at USACERL to meet the requirements for data management of large computer software systems which must utilize data originating from many different sources and in several different storage formats. Results produced by the experimental software have exceeded expectations. Microcomputer response times have been very acceptable, even on very computer intensive operations and very large databases. The results of this research provide insights into the question of how to develop large computer systems such as the TCMS which must have a totally integrated database to achieve the highest level of effectiveness. The results also show that microcomputer class machines are acceptable vehicles for this class of computer software system provided the machines meet the minimum requirements stated above. PSE software is being developed by the Army for the Army. The advantage implicit in this statement is that the software can be expanded with minimal cost and distributed without distribution fees paid to a commercial vendor. As new requirements are identified the developers at USACERL and the software maintainers at Huntsville Division of the Corps of Engineers can implement these requirements without the need to work with a third party.

Experimental Development of a
Microwave Vlasov Mode Convertor (U)

*Robert K. Dahlstrom, Mr.
Brian G. Ruth, Mr.
Larry J. Hadwin, Mr.
Louis F. Libelo, Dr.
Harry Diamond Laboratories
Adelphi, MD 20783-1197

1. INTRODUCTION

Many of the high-power microwave (HPM) sources in an advanced state
of development (e.g., the relativistic backward-wave oscillator (BWO),
the gyrotron, and the virtual cathode oscillator (vircator)) generate
power in cylindrically symmetric transverse electric (TE) or transverse
magnetic (TM) modes. These modes are not suitable for directly driving
conventional aperture antennas because they produce a null on boresight.
Use with a reflector would result in excessive sidelobe generation and
gain reduction. One method of optimizing antenna performance in
conjunction with these HPM sources is the use of a Vlasov mode convertor,
a device that converts circular TM or TE modes to a good approximation of
a plane-parallel linearly polarized Gaussian-profile beam.[1] In the past,
however, the operating range of Vlasov mode convertors has been
restricted to quasioptical frequencies of 60 GHz or above (used mainly
for plasma heating) and to TE modes. In theory, this type of mode
convertor should scale down to microwave frequencies. When used at high
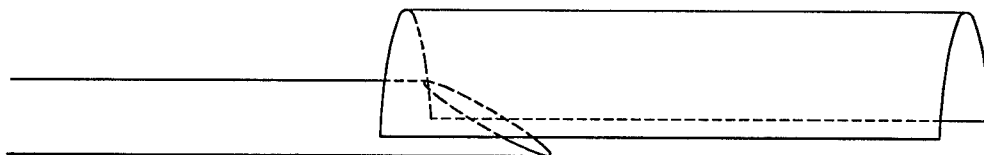

Figure 1. Vlasov mode convertor with a bevel-cut aperture
and one parabolic cylinder reflector.

power levels in conjunction with electron-beam-driven microwave sources, the Vlasov mode convertor could prove useful as part of a ground-to-space power transmission system or a susceptibility and hardening test system.

Figure 1 shows the mode convertor. It consists of a circular waveguide with a shaped open-end aperture (Vlasov radiator) coupled with one or more focussing reflectors (only one is indicated). The aperture directs the microwave energy in a given distribution over the surface of the reflector, which in turn redirects the energy into a beam with a parallel wave front, or some other directional characteristic. In the design of such a radiator the configuration of the radiator aperture must be optimized for the waveguide mode to be efficiently radiated.

In this paper, we measure the radiation patterns from three different Vlasov radiators, which have beveled apertures cut at angles of 10°, 15°, and 20° relative to the waveguide axis and compare them to previous measurements taken on five other radiators with bevel angles ranging from 30° to 60°. A step-cut aperture, as originally described by Vlasov,[1] has sharp corners and therefore may suffer from electrical breakdown when used at high-power levels. The beveled cut, suggested by Nakajima,[2] avoids these sharp corners, and may be more suitable for high-power applications. We also measure the phase center (which will be described in greater detail in section 2.2) of the microwave energy at the aperture of the 30° bevel. This information is needed to determine the physical characteristics of the reflector component of the Vlasov system. The radiation pattern of such a Vlasov radiator coupled with a parabolic cylinder reflector is presented. In addition, the apertures
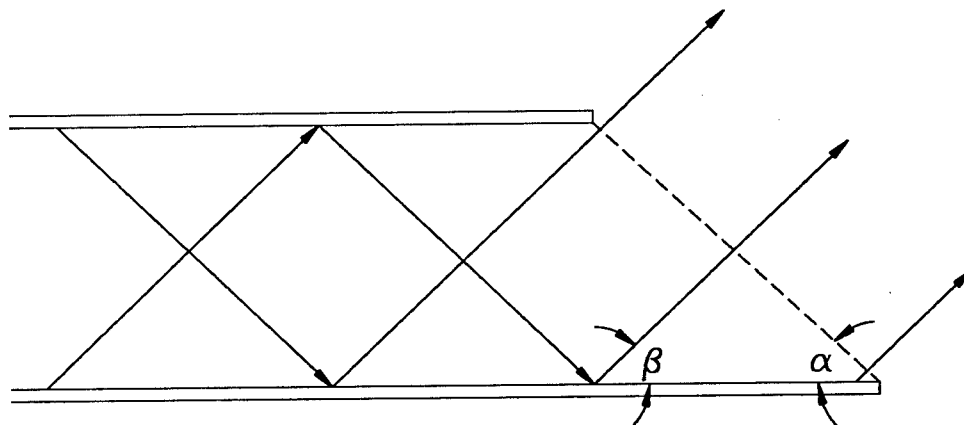

Figure 2. Propagating quasioptical rays launched from a bevel-cut rectangular waveguide.

with bevel angles less than 30° are evaluated for use without a reflector and their measured performance is described.

2. BASIC THEORY

2.1 Vlasov Radiator

To a first approximation, the operating theory of a quasioptical Vlasov mode convertor is derived from geometrical optics.[3] Let us assume, for simplicity's sake, that the waves inside a circular waveguide behave in a similar manner to rays propagating in rectangular waveguide at an angle β relative to the broad wall of the waveguide, as shown in cross-section in figure 2. The radiating aperture of this waveguide as seen in figure 2 can be compared to circular waveguide cut at an angle α relative to its external surfaces. Then, for an efficient radiator, the aperture should be large enough to allow all the rays representing the uniform phase front of the waveguide mode to exit through the aperture; that is, $\alpha \leq \beta$, so that all the propagating rays are launched at the same angle in the same direction. If $\beta < \alpha$, then some of the rays are launched in a different direction and do not contribute to the desired beam. Therefore, if we choose to set $\alpha \leq \beta$, then the aperture angle, α, should be approximated by

$$\cos \alpha \geq \cos \beta = c/v_{ph}, \tag{1}$$

where v_{ph} is the magnitude of the phase velocity of the waveguide mode and c is the speed of light. The quantity v_{ph} is also related to c by

$$v_{ph} = c(\lambda_g/\lambda_0), \tag{2}$$

where λ_g is the waveguide wavelength and λ_0 is the free-space wavelength. These wavelengths can in turn be expressed for a circular waveguide by

$$\lambda_g = 2\pi[(\omega/c)^2 - (p_{nm}/a)^2]^{-1/2} \tag{3}$$

and

$$\lambda_0 = 2\pi c/\omega, \tag{4}$$

where ω is the circular frequency, a is the inner radius of the waveguide, and p_{nm} is the mth root of the Bessel function $J_n(p_{nm})$. Substituting equations (3) and (4) into (2) yields

$$v_{ph} = c[1 - (p_{nm}c/(\omega a))^2]^{-1/2}. \tag{5}$$

Further substitution of (5) into (1) yields

$$\cos^2\alpha \geq \cos^2\beta = 1 - (p_{nm}c/(\omega a))^2. \tag{6}$$

Finally, applying the trigonometric relation $\sin^2\alpha + \cos^2\alpha = 1$ to equation (6) yields

$$\sin\alpha \leq \sin\beta = p_{nm}c/(\omega a). \tag{7}$$

The above relation is only approximately correct for circular waveguide, since this analysis is an extrapolation of the analysis for rectangular waveguide where there is no azimuthal field component.[4] Also, this analysis does not take into account complications such as diffraction at the waveguide aperture.

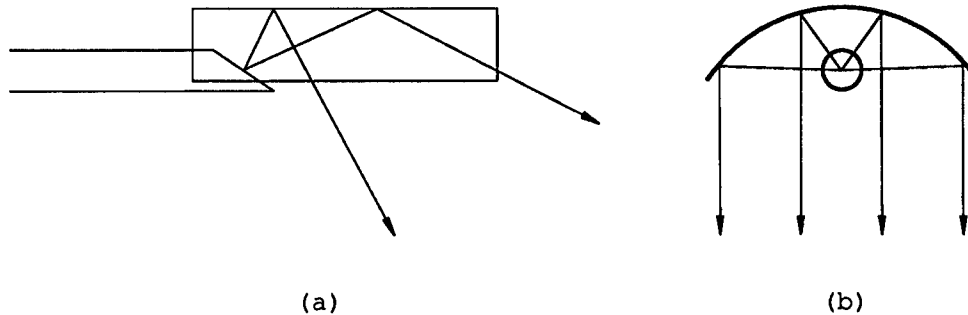(a)   (b)

Figure 3. Interaction of quasioptical rays with a parabolic cylinder reflector. (a) Broadside view. (b) End-fire view.

2.2 Vlasov Reflector

Once the aperture dimensions of a Vlasov radiator are selected and the radiation pattern characteristics measured, the reflector can be designed. The needed radiation pattern characteristics include θ and ϕ-plane radiation patterns and sufficient phase measurements to locate the phase center if it exists. This phase center is an imaginary point from which, in the far-field radiation limit, all the waves appear to originate. The parabolic cylinder reflector is traditionally placed close to the aperture as shown in figure 3. The parabolic surface is shown in figure 3(a) with its focal point aligned with the measured phase center of the aperture to focus the broad beam in ϕ before it is allowed to disperse. Figure 3(b) shows the θ directed energy deflected in beam angle, but with its beam shape characteristics virtually unchanged. The

size of the reflector in both planes is chosen to sufficiently capture the beam based on the measured radiation patterns. For a high-efficiency design, the reflector would be extended out far enough to capture radiation from the aperture that is 20 dB down from the peak value or more.
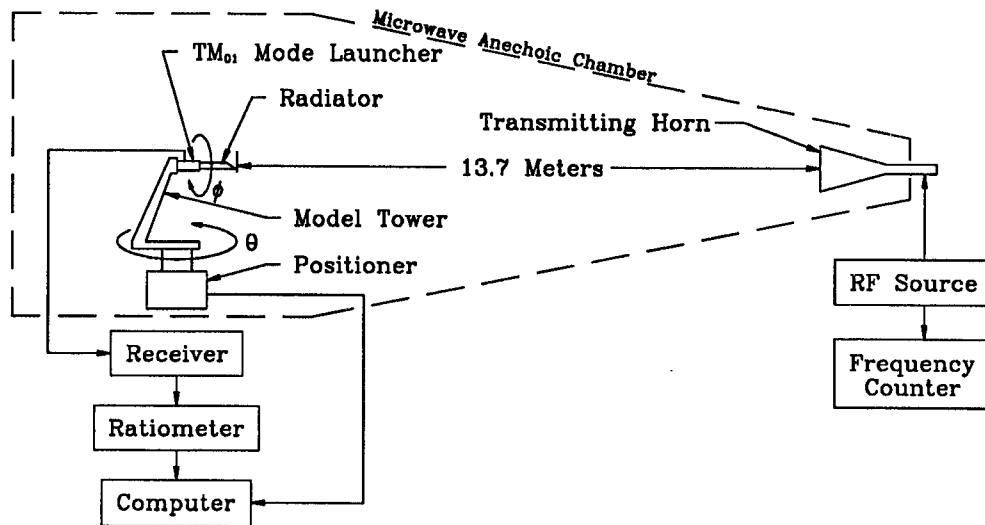

Figure 4. Antenna pattern measurement system.

3. MEASUREMENTS

3.1 Antenna Pattern Measurement System

Figure 4 shows the antenna pattern measurement system used to record the directional characteristics of the mode convertor designs. The Vlasov mode convertor under test is connected to the output end of a circular TM_{01} mode launcher. This mode launcher/mode convertor combination is mounted on a model tower. The head of the tower provides roll plane rotation about the azimuthal angle, ϕ. The whole tower is mounted on a pedestal, which can rotate about a vertical axis to vary the polar angle, θ.

Although the mode convertor is expected to be used as a transmitter, reciprocity allows us, for convenience, to measure the radiation pattern of a mode convertor while it is being used as a receiver. Illumination is provided by a horizontally polarized horn antenna driven by a cw source at 8.6 GHz. The signal intercepted by the

(a)

(b)
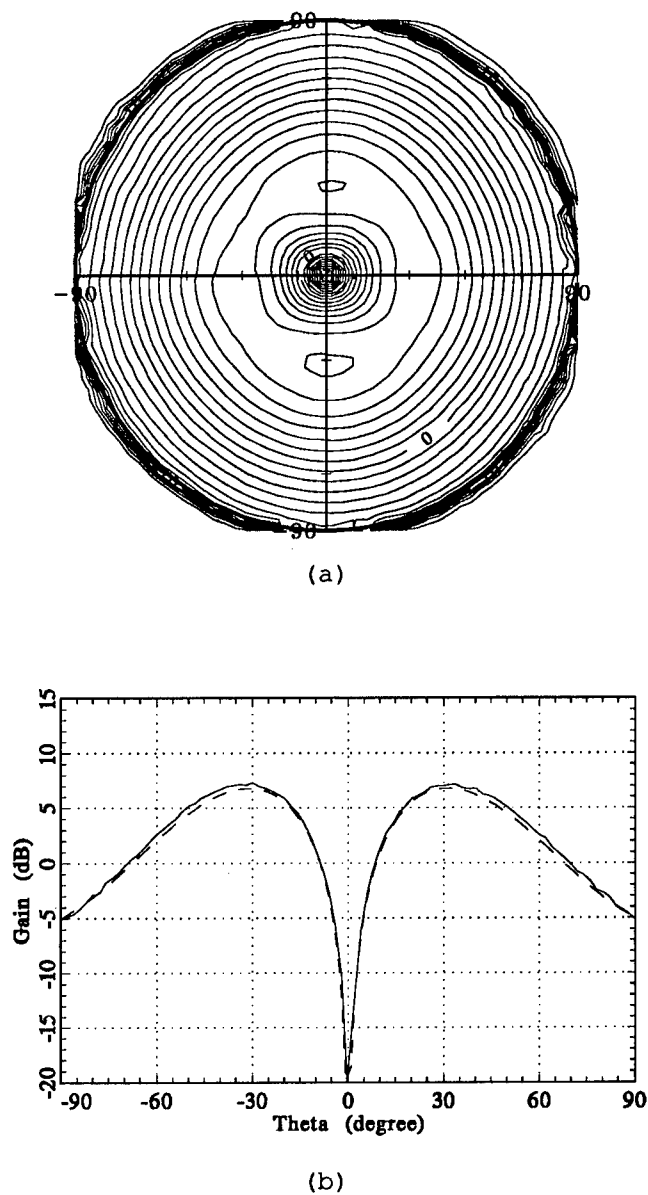
Figure 5. Measured antenna pattern from a straight-cut circular
waveguide radiating in the TM_{01} mode. (a) Contour plot of gain as
a function of ϕ and θ; contour lines are shown every dB, the
azimuthal angle ϕ varies from 0° to 360°, and the elevation angle θ
varies from 0° in the center to 90° at the edge. (b) Elevation cut
showing gain as a function of θ for $\phi = 0°$. Measured data is
plotted as a solid line, calculated data as a dashed line.
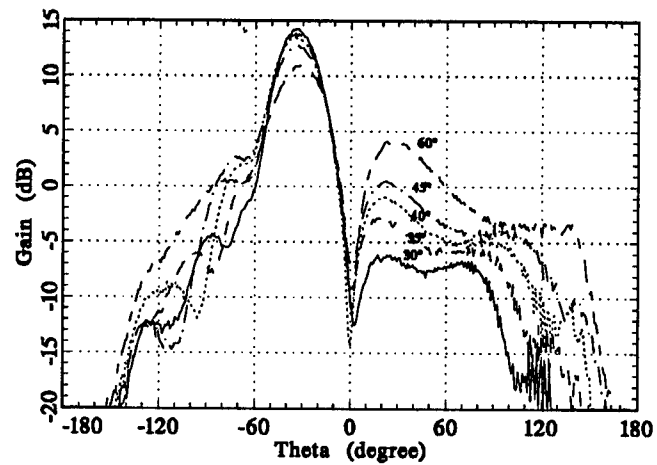
(a)
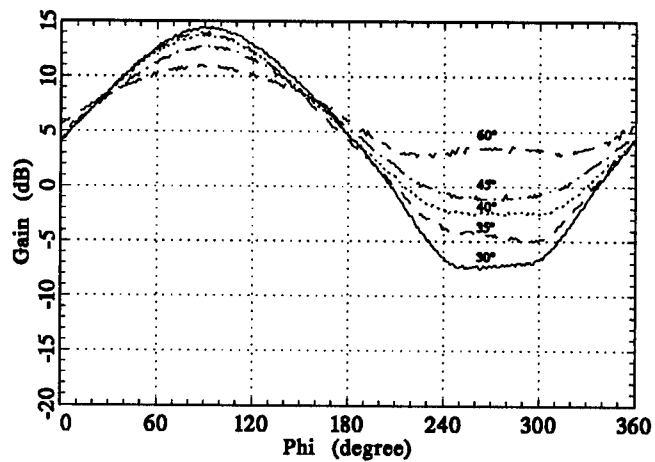
(b)

Figure 6. Previously measured antenna patterns of five bevel-cut
apertures with various bevel angles. (a) Elevation cuts showing
gain as a function of θ for φ = 90°. (b) Azimuthal cuts showing
gain as a function of φ at the beam peaks in θ (32° to 35°).

(a)

(b)

Figure 7. Measured antenna patterns of three new bevel-cut apertures
with various bevel angles. (a) Elevation cuts showing gain as a
function of θ for $\phi = 90°$. (b) Azimuthal cuts showing gain as a
function of ϕ at the beam peaks in θ ($32°$ to $35°$).
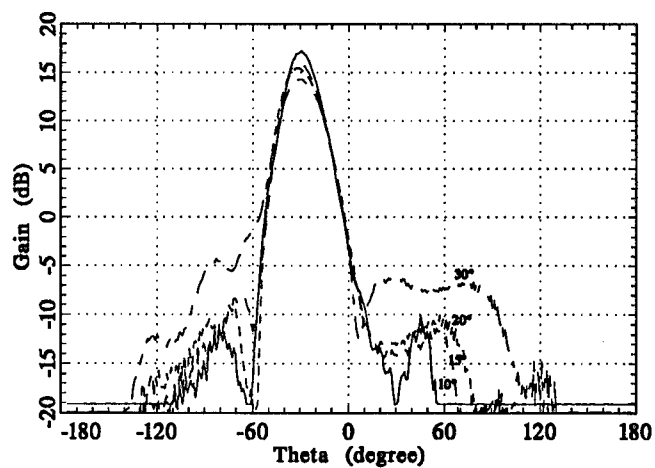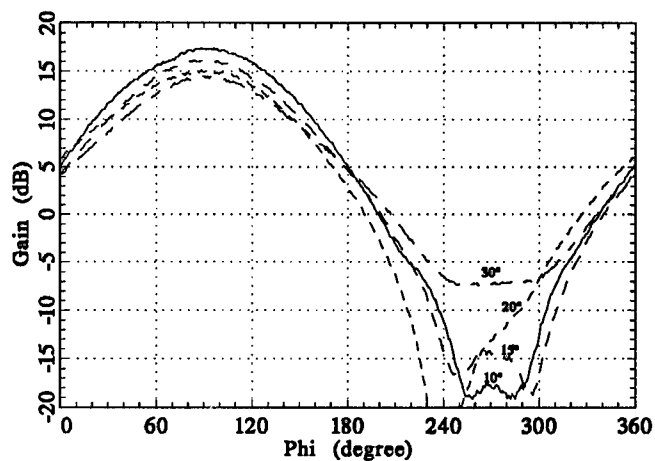
mode convertor is back-converted to a circular TM_{01} waveguide mode, which then enters the mode launcher and is further back-converted to two in-phase rectangular TE_{10} waves. These waves are added together in-phase by an X-band magic tee, mixed down to 1 kHz by the receiver, and then converted to a dc voltage that is proportional to received power in dB by the ratiometer. This voltage, and the angles ϕ and θ from the model tower, are supplied to a computer for storage on disk and later processing.

3.2 TM_{01} Mode Verification

In its final operating stage, the Vlasov mode convertor will be driven by a high-power backward-wave oscillator (BWO) with an operational frequency of 8.4 to 8.8 GHz. Therefore, for the Vlasov radiator, we choose a design frequency of 8.6 GHz. If we use a standard-size stainless-steel circular pipe of inner radius a = 2.38 cm for the radiator, then five circular modes will be able to propagate down the guide (TE_{11}, TM_{01}, TE_{21}, TM_{11}, and TE_{01}). However, the symmetric driving of the TM_{01} mode launcher should preclude both asymmetric modes and modes without an E_z component. These restrictions remove all but the TM_{01} mode from the guide. This radius waveguide is used for its ability to carry high levels of microwave power.

To determine that we do indeed have an isolated TM_{01} mode in the waveguide, we first construct a test radiator with a straight-cut aperture (α = 90°) and measure the radiation pattern. This pattern is shown in Figures 5(a) and 5(b). Figure 5(a) illustrates the azimuthal symmetry of the radiation. In Figure 5(b), an elevation cut taken at ϕ = 0 is plotted against the calculated far-field radiation pattern from a circular waveguide radiating in the TM_{01} mode. From the azimuthal symmetry in Figure 5(a) and the reasonable agreement between theory and measurement in Figure 5(b), we can assume a single TM_{01} mode in the waveguide.

3.3 Vlasov Radiator Design and Gain Measurements

Next, we design the Vlasov radiator. If p_{01} = 2.405 (for a TM_{01} mode), then, from equation (7) in section 2.1, $\sin \alpha \leq 0.561$ or $\alpha \leq \sin^{-1}(0.561)$ = 34.1°. Previously, the radiation patterns from five different radiators, each 40 cm in length, with bevel angles of a = 30°, 35°, 40°, 45°, and 60° were measured and reported by Ruth.[4] These measurements are plotted in figures 6(a) and 6(b). In these comparative plots, we see the formation of a beam increasing in gain as a function of decreasing α. This beam is well-focused in the θ-plane but relatively broad in the azimuthal (ϕ) direction. The α = 30° bevel is just below the calculated upper limit for α, and is thus the best performer of the
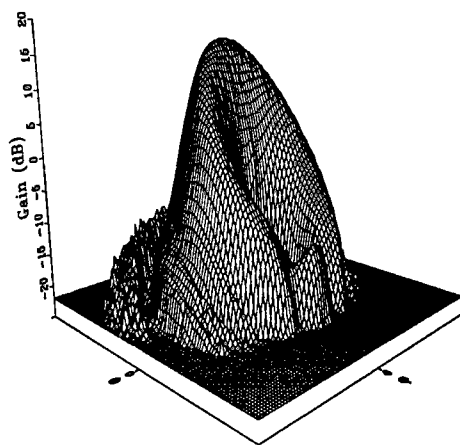
(a)
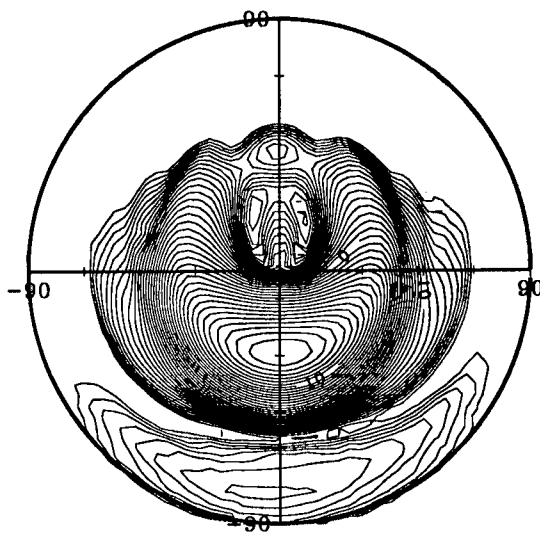
(b)

Figure 8. (a) Three-dimensional plot and (b) contour plot, showing gain as a function of ϕ and θ, of the measured antenna pattern of the α = 10° bevel-cut aperture.
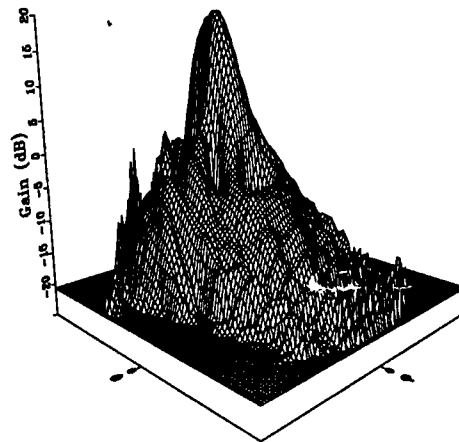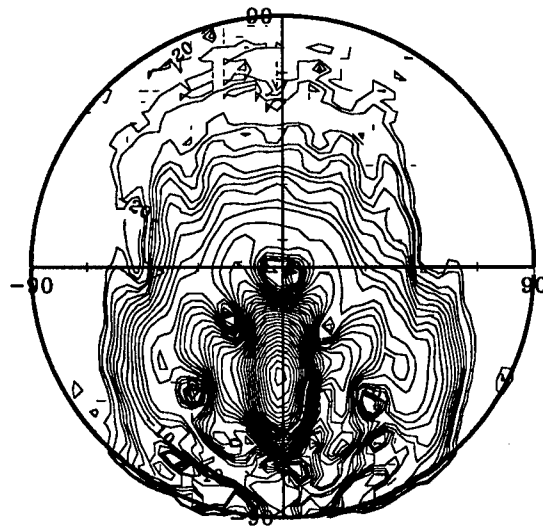
(a)

(b)

Figure 9. (a) Three-dimensional plot and (b) contour plot, showing gain
as a function of ϕ and θ, of the measured antenna pattern of the α
= 30° bevel-cut aperture coupled with a parabolic cylinder
reflector.

five radiators tested. To investigate the behavior of waveguide apertures where α < 30°, three new radiators were fabricated with bevel angles of α = 10°, 15°, and 20°. These measurements, along with the α = 30° bevel for comparison purposes, are plotted in figures 7(a) and 7(b). The trend observed in the first five radiators is continued in the latter three radiators; that is, the gain of the main beam increases and the width of the beam narrows (in the θ-plane) as the bevel angle is decreased. It should be noted that the beam no longer narrows in the φ-plane and that the sidelobes appear to fall to a residual level at a bevel angle between 20° and 30°. All the measured three-dimensional antenna patterns are remarkably similar to one another in that a horseshoe-like beam is formed about 32° to 35° above the boresight power null. Thus, in terms of high gain and narrow main beam, the α = 10° radiator was the best performer. The measured three-dimensional antenna pattern of the α = 10° radiator is plotted in figure 8.

3.4 Single-Reflector Configuration

As shown in figures 6(a) and 7(a), the bevel-cut radiator emits energy in a relatively narrow beam in θ and, in figures 6(b) and 7(b), over a broad range of angles in φ. A reflector to be used with the 30° aperture was designed as described in section 2.2 to narrow the beam further in φ. Referring to the measured radiation pattern, capturing energy down 10 dB below maximum requires a reflector extending at least 180° around the radiator. The phase center of the radiator, found from measured phase patterns, is located on-axis 7 cm beyond the center of the aperture. The length of a reflector needed to intercept emissions down to the 10 dB level was found to be at least 8 inches at 8.6 GHz.

The measured performance of this configuration is given in figure 9. Comparing the performance of the 30° bevel aperture with and without the reflector, the half-power width of the beam is narrowed from 87° to 16° in φ, and the gain is increased from 14 to 20 dB. The addition of a reflector to the 10° bevel is being considered, but the increased reflector blockage of its longer aperture is a concern.

4. CONCLUSIONS

Among the eight Vlasov radiators tested, we found consistently that peak radiation gain levels were an inverse function of bevel angle α. As the bevel angle descends from α = 90° to α = 10°, the peak gain level increases from 7.3 to 17 dB. The increase in directivity is attained predominantly by a narrowing of the beam width in the θ-plane for bevel angles less than 30° and by a narrowing in both planes for the larger bevel angles. The highest gain of 20 dB was attained with the use of a reflector, but the amount of increase in gain may be insufficient to

justify the added physical complexity of this configuration. The quasioptical analysis of the Vlasov radiator, with its implied approximations, failed to explain the exact mechanism behind the mode-converting beveled aperture for the configurations measured herein. However, this analysis did predict the presence of a transition region which appears in the measured ϕ-plane beamwidth characteristics and in the measured level of the sidelobes at a bevel angle between 20° and 30°. A more rigorous approach to the analysis of the Vlasov radiator is presently being carried out in collaboration with Rensselaer Polytechnic Institute.

(1)   S. N. Vlasov and I. M. Orlova, "Quasioptical transformer which transforms the waves in a waveguide having a circular cross section into a highly directional wave beam," Radiofizika, Vol. 17, No. 1, pp. 148-154, January 1974.

(2)   O. Wada and M. Nakajima, "Quasi-Optical Reflector Antennas for High Power Millimeter Waves," Proceedings of the EC6-Joint Workshop on ECE and ECRH, Oxford, Sept. 1987.

(3)   R. P. Feynman, Leighton, R. B., and Sands, M., The Feynman Lectures on Physics, Addison-Wesley Pub. Co. (1964), Vol. 2, pp. 24-11, 24-12.

(4)   B. G. Ruth, R. K. Dahlstrom, and C. D. Schlesiger, "Antenna Measurements on the Radiator Component of an X-Band Vlasov Mode Convertor," Intl. Journal IR and MM Waves, Vol. 10, No. 7, July 1989.

Predicted Flight Performance of Base Bleed Projectiles (U)

James E. Danberg* and Charles J. Nietubicz
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland 21005-5066

I. INTRODUCTION

The primary objective of this work is the development of an engineering analysis model for the M864 projectile flight performance. A schematic of the M864 including the base bleed propellant in the afterbody is shown in Figure 1. The engineering model considers all three elements of the problem; namely: gas generation, effect of bleed on aerodynamic drag, and a trajectory model. Because each of the above elements is time dependent, all three elements must be solved simultaneously. This is in contrast with a conventional projectile where the aerodynamic coefficients can be determined independent of any specific trajectory. The analysis is

Figure 1: Schematic of M864

based on much of the existing experimental and numerical data obtained as part of the M864 program. However, the techniques developed are easily adapted to other designs.
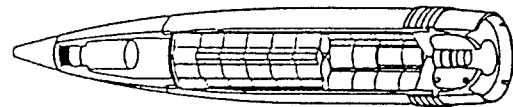
A secondary objective is to determine how the numerical solutions to the Navier-Stokes equation can be applied to the problem. At present these computations do not account for the time dependent nature or real gas properties of the gas generator and the changing free-stream conditions along the trajectory. Thus, steady state solutions are obtained for specific free-stream conditions with equivalent air mass flow and temperature imposed as boundary conditions at the exit of the gas generator. The results of these computations are used for correlations of base pressure as a function of injected mass flow and temperature.

A broad survey of the aerodynamics of base bleed/burn was performed for a meet-

ing in 1974.[1] The Proceedings of that meeting contained an extensive bibliography of work reported to that time. Since 1974, much new work has been carried out, particularly in developing and testing practical projectile designs using base bleed, such as the U.S. Army's M864. References to the current international base bleed projectile literature are contained in the Proceedings of the First International Symposium on Base Bleed.[2]

II. GAS GENERATOR

The M864 gas generator is housed in the boattail afterbody. It consists of two identical solid propellant grains. These grains provide an inner cylindrical burning surface and four planer surfaces separated by a 3 mm slot. Miller and Holmes[3] measured the HTPB propellant burning rate over a range of sub-atmospheric pressures using a strand burning technique. The strand burning rate, $\dot{r}_o$, was found to be $\dot{r}_o = 0.00009132 P_c^{0.6655}$ where P_c is in bars and $\dot{r}_o$ is in m/s. The burning rate on the cylindrical surface of M864 propellant, $\dot{r}_c$, and on the slot surface, $\dot{r}_s$, are considered uniform over the surfaces but may possibly be different. Thus the mass generated by the solid propellant can be written:

$$\dot{m} = \rho_s(\dot{r}_c A_c + \dot{r}_s A_s) \tag{1}$$

1. Tests in a Ground Based Spin Fixture

Gas generator tests have been performed where an actual M864 motor was mounted on a spin fixture.[4] Chamber temperature, chamber pressure and burn time were measured for various rates of spin. The chamber pressure measurements are shown in Figure 2. The measured chamber pressure-time histories can be interpreted as indicating that there is a difference between the burning rate on the cylindrical surface and on the surface of the slot. The chamber pressure is directly an indication of the exit mass flow. If the burning rates are equal on the two surfaces then the chamber pressure should decrease monotonically with time, after an initial transient, because of the shape of the grains. As shown in Figure 2, at high spin rates the pressure initially increases and then falls off, indicating that the cylindrical surface is more important in determining the mass flow, and thus chamber pressure than the monotonically varying slot contribution.

The mass flow, $\dot{m}$, is assumed subsonic but described by the following compressible

flow equation:

$$\dot{m} = K \sqrt{\frac{\gamma}{(\gamma - 1)} \left[(\frac{P_c}{P_b})^{\frac{\gamma - 1}{\gamma}} - 1 \right]}. \tag{2}$$

where P_c is the chamber pressure ,P_b is the external pressure and γ is the ratio of specific heats of the exhaust products (γ=1.25). The factor K can be written:

$$K \equiv C_v A_j \sqrt{2\rho_j P_b}. \tag{3}$$

where C_v, A_j, and ρ_j are the discharge coefficient, exit area and, exit density, respectively. Since each is approximately constant, K may also be taken as constant in these laboratory tests. In the zero spin case, the regression rates are presumed to be the same on both surfaces, and approximately equal to the strand burning $\dot{r}_o$, since the chamber pressure history is basically monotonically decreasing with time. The zero-spin chamber pressure distribution can be used to determine the constant K, which according to Kayser's. data,[4] gives K=0.189 kg/s.

2. Effect of Spin

When spinning, the cylindrical and slot surfaces are assumed to have different regression rates which are determined as follows. The radial thickness of the grain is the smallest dimension and thus the total burn-time determines the burn rate on the cylindrical surface. The mass flow, Equation (1), is integrated over time for the known amount of propellant using different $\dot{r}_c$ values until the calculated burn-time matches the measured time.
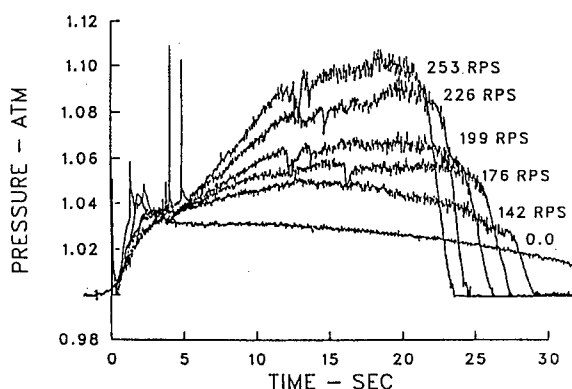

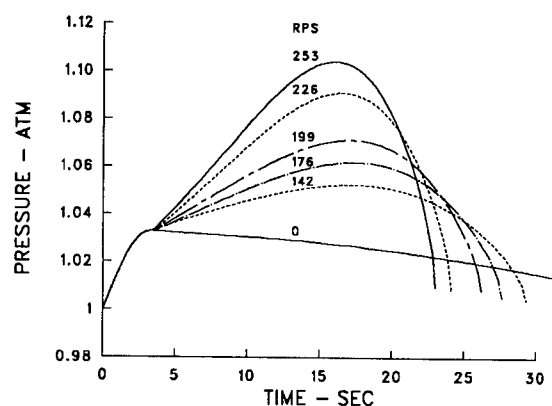Figure 2: Experimental Chamber Pressures[4]

Figure 3: Computed Chamber Pressure

Next, it is observed (see Figure 2) that there is an initial transient which is relatively independent of spin. The pressure is 3.3 per cent higher than ambient at 3.3 sec-

onds after ignition for all spin rates. After that point, the chamber pressures diverge, depending on spin. The regression rates were computed during the initial transient phase assuming they were identical but time varying.

Finally, the slot contribution affects the magnitude of $\dot{m}$ but it does not affect the burnout time. Conditions at the end of the initial transient phase define the burning rate of the slots. This is equivalent to assuming that the mass flux from the gas generator is independent of spin at the end of the initial transient phase.

The results are shown in Figure 3 where the predicted chamber pressure is plotted against time for the various spin rates of the spin fixture tests. This figure can be compared to Figure 2. Very good agreement is obtained, particularly during the early period, when the effect of the different spin rates is being observed. The most important result for subsequent calculations of the flight case is Figure 4 which shows $\dot{r}_c$ and $\dot{r}_s$, normalized by the stand burning rate, as a function of spin.
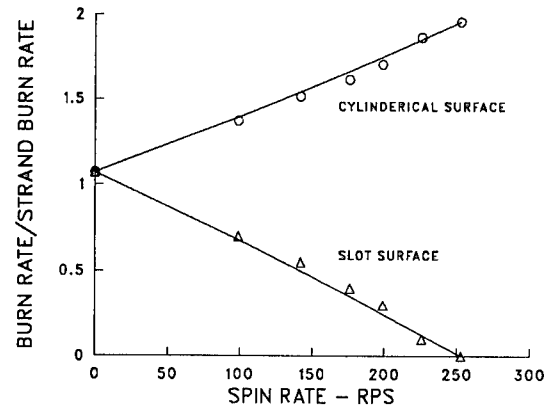
Figure 4: Effect of Spin on Burn Rate

3. Application to Flight Test

In applying the above analysis to the flight situation, it is assumed that the exit pressure, P_b, becomes the base pressure and:

$$K_{flight} = K \frac{P_b}{P_{sl}},$$

where P_{sl} is sea level pressure. The mass flow generation formula, equation (1), remains the same. Equations (1) and (2), may be considered, at each point in time, as two equations with four unknowns, $\dot{m}$, P_b/P_∞, P_c/P_∞, P_∞/P_{sl}. Along the trajectory M_∞, and P_∞/P_{sl} would be known, and aerodynamic theory gives a third relation between $\dot{m}$ and P_b/P_{sl} (also involving M_∞ and P_∞/P_{sl}) which determines the solution at each instant in time.

The solid propellant fuel is relatively difficult to ignite and, therefore, the gun gases are supplemented by an igniter which consists of two 6.2 gram cylindrical pellets of magnesium-teflon. The igniter is consumed in two seconds and contributes significantly to the mass flow in the first seconds of flight.

414

III. AERODYNAMICS OF BASE INJECTION

It is customary when considering base bleed or burn to define a nondimensional injection parameter, I:

$$I = \frac{\dot{m}}{\rho_\infty u_\infty A_b} \tag{4}$$

In the M864 case, the average value of this parameter is about 0.002 at launch and increases with decreasing Mach number and increasing altitude.

Most supersonic wind tunnel experiments on this subject are concerned with cold gas injection into the base region at relatively large values of I. Figure 5 shows the results of one such experiment by Kayser.[5] He investigated the effect of the injector geometry (i.e, per cent open area of the base) on the base pressure. The results are for a Mach number of 3.0 with cold air injected into the model wake. Note that there is a strong dependence on injector open area at large values of I. However, all the curves originate at the same pressure and increase linearly at the same slope for values of I less than 0.005. This suggests that, at small injection rates, a linear increase in base pressure can be assumed, although as the injection rate gets larger such an approximation may over-estimate the beneficial effects.
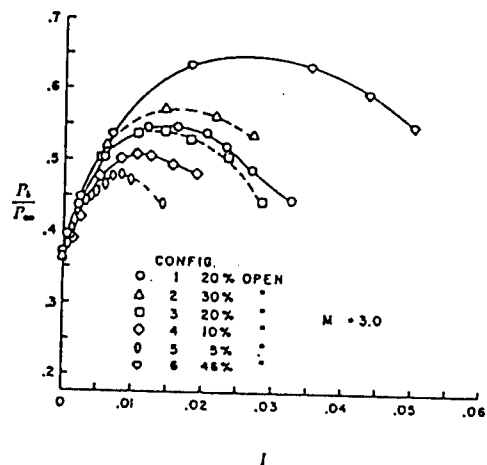
Figure 5: Measured[5] Base Pressure vs I

1. Linear Effect of Injection on Base Pressure

As a preliminary approach, it is assumed that the mass injection rate is at sufficiently low subsonic velocities that the injection effect on base pressure is nearly linear, i.e.,

$$\frac{P_b}{P_\infty} = \left(\frac{P_b}{P_\infty}\right)_{I=0} + \left[\frac{d(P_b/P_\infty)}{dI}\right]_{I=0} I. \tag{5}$$

A number of reports on wind tunnel measurements of base pressure with injection have been examined. Figure 6 shows the results of determining $[\frac{d(P_b/P_\infty)}{dI}]_{I=0}$ as

415

a function of the free stream Mach number. The open symbols represent wind tunnel experiments in which ambient room temperature air was injected. Although these data correlate reasonably well with Mach number, they represent a special case of low temperature injection into a low temperature wake. (stagnation temperature of the wind tunnel flow is estimated also to be ambient room temperature) Also included in Figure 6 are three points (filled symbols) computed by Nietubicz.[6] The magnitude of the numerical results agree quite well with the wind tunnel data.
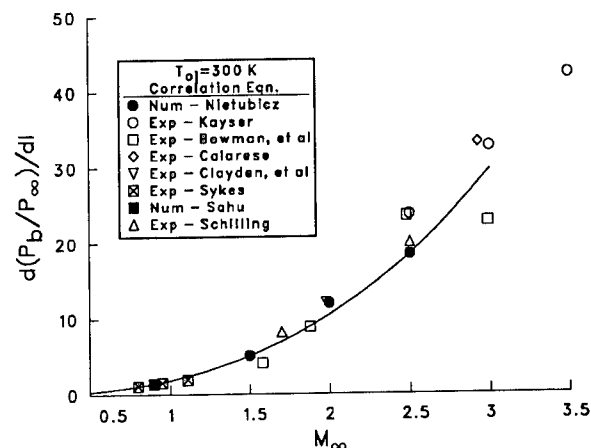
Figure 6: $\frac{d[P_b/P_\infty]}{dI}$ vs M_∞, $T_{o_j}=300K$

Other numerical computations[7] of the near wake flow with injection have been performed to broaden our data base for these kinds of flows. These computations, which involve solution of the perfect gas Navier-Stokes equations, are discussed in the following sections.

2. Navier Stokes Computational Technique

The time-dependent, thin-layer, Navier-Stokes equations are solved numerically to obtain a solution to this problem. The numerical technique used is an implicit finite difference scheme. The steady flow, which is the desired result, is obtained in a time asymptotic fashion. The time-dependent, thin-layer, Navier-Stokes equations written in strong conservation law form for the axisymmetric formulation[8] are:

$$\frac{\partial \hat{q}}{\partial \tau} + \frac{\partial \hat{E}}{\partial \xi} + \frac{\partial \hat{G}}{\partial \zeta} + \hat{H} = \frac{1}{Re}\frac{\partial \hat{S}}{\partial \zeta} \tag{6}$$

The general coordinate transformations are defined as:

$$\xi = \xi(x,y,z) \text{ - is the longitudinal coordinate}$$
$$\zeta = \zeta(x,y,z) \text{ - is the near normal coordinate}$$
$$\tau = t \text{ - is the time}$$

The vector $\hat{q}$ contains the dependent variables $[\rho, \rho u, \rho v, \rho w, e]$ and the flux vectors $\hat{E}, \hat{G}$ contain terms which arise from the conservation of mass, momentum and energy. The source vector $\hat{H}$, in Equation (6), contains terms which result from an analytic

416

determination of the circumferential flux vector given the assumption of axisymmetric flow and constant angular velocity.[8] The viscous terms are contained in the vector $\hat{S}$ which is seen to have variation in the ζ direction only. This is representative of the thin-layer approximation.

The numerical algorithm used for the solution of Equation (6) is the Beam and Warming[9] implicit, approximately factored, finite-difference scheme which uses central differencing in both ξ and ζ directions. Code improvements have been made to include a variable time step, numerical smoothing based on local solution gradients and code vectorization.[10] The Beam-Warming implicit algorithm has been used in various applications for the equations in general curvilinear coordinates. The algorithm is first-order accurate in time and second- or fourth-order accurate in space. Central difference operators are employed; the algorithm produces a block tridiagonal system for each space coordinate. The main computational work is contained in the solution of these block tridiagonal systems of equations.

The axisymmetric code that solves Equation (6) uses a unique flow field segmentation procedure to compute the full flow field over a projectile or a missile including the base region. Although the thin-layer approximation is not valid in the wake region, this segmentation procedure allows for turbulent quantities to be determined across the shear layer.[11] This has provided reasonable accuracy for supersonic flows. For the computation of turbulent flows, the two-layer algebraic, Baldwin-Lomax,[12] turbulence model is used over the projectile body.

3. Boundary Conditions

The outer computational boundary, where free stream conditions were imposed, was kept at approximately 12 body diameters from the body surface. At the downstream boundary extrapolated outflow conditions were used. Viscous no slip boundary conditions were imposed on all solid surfaces. The normal momentum equation was solved at the wall for p_n, the normal pressure gradient. The free flight wall temperature was specified. Using the time lagged pressure, the density at the wall is determined.

The boundary conditions on the exit of the gas generator follow from the computed pressure and flow field variables at each time step. The bleed boundary conditions are thus lagged in time. The bleed conditions $(pu)_j$, and T_{o_j} are specified while p_j at the gas generator exit is extrapolated from the exterior flow. By specifying the mass injection rate, I, the bleed Mach number, M_j, can be directly solved by:

$$M_j^2 = \frac{2-\gamma}{\gamma-1} + \sqrt{1 + 2(\gamma-1)\frac{T_{o_j} I M_\infty \rho_\infty A_b^2}{T_\infty \rho_j A j}} \tag{7}$$

where A_b and A_j are the area of the base and bleed opening respectively. Using isentropic relations the bleed stagnation pressure p_{o_j} and static temperature T_j can be calculated. Given the temperature and pressure the density is computed and the boundary conditions for the next time step are obtained.

4. Model Geometry and Computational Grid

The computational model of the M864 is shown in Figure 7. Projectile features which have not been modeled exactly are the fuse meplat and the rotating band. The rotating band was eliminated. The meplat was modeled as a hemisphere cap. The model nose is a combination of ogive and conical sections. The cylindrical section is slightly undercut, and the boattail length is 0.524 calibers with a 3° angle. The shape of the base is shown as a dotted line on the boattail. The base bleed motor is located internally with the exhaust port centered on the model axis.
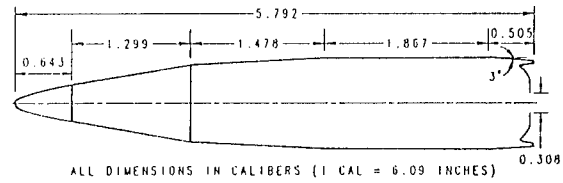
Figure 7: Computational Model

The solution technique requires the discretization of the entire region of interest into a computational grid. The grid outer boundary has been placed at 1.5 body lengths upstream and surrounding the projectile. The downstream boundary was placed at 2 body lengths. Since the calculations include subsonic/transonic regimes, the computational boundaries must extend out beyond the influence of the body.
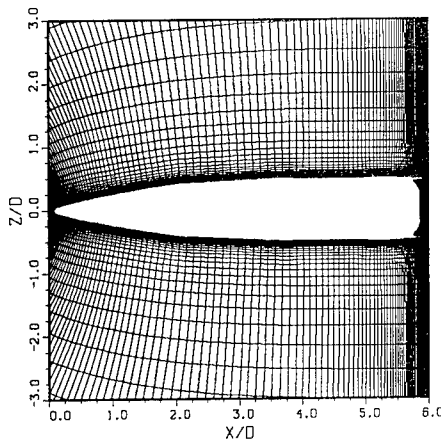
Figure 8: Grid Near M864 Model
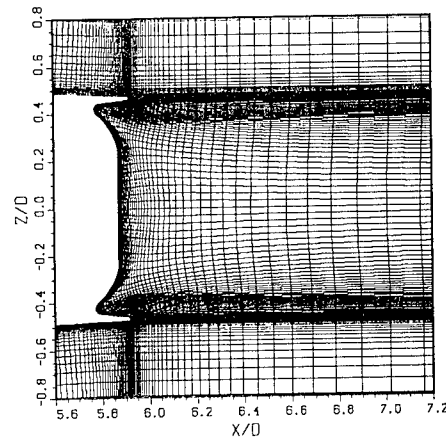
Figure 9: M864 Dome Base Grid

418

Figure 8 shows the grid generated over the projectile body including the base. It consists of 265 longitudinal points and 60 points in the normal direction. This is broken down into two sections: a body region, and a base region. The surface points for each region are selected using an interactive design program. Each grid section is then computed separately using a hyperbolic grid generation program.[13] Figure 9 shows an expanded view for the dome base grid. The configuration for the concave dome base required an increase in the smoothing values used by the hyperbolic grid generator as well as the addition of a grid cell averaging technique.

5. Computational Results and Correlation

Computations have been performed for the M864 dome base with the inclusion of base bleed. The computations to date have been for cold and hot mass flow, with values of the mass injection parameter,I, up to 0.04. Although the actual mass flow rate varies over the course of the trajectory, the present calculations were run to see the effect on drag across a Mach number range.

As an attempt to model the effects of burning, computations were run with hot mass addition. The results shown in Figures 10 and 11 are particle traces in the
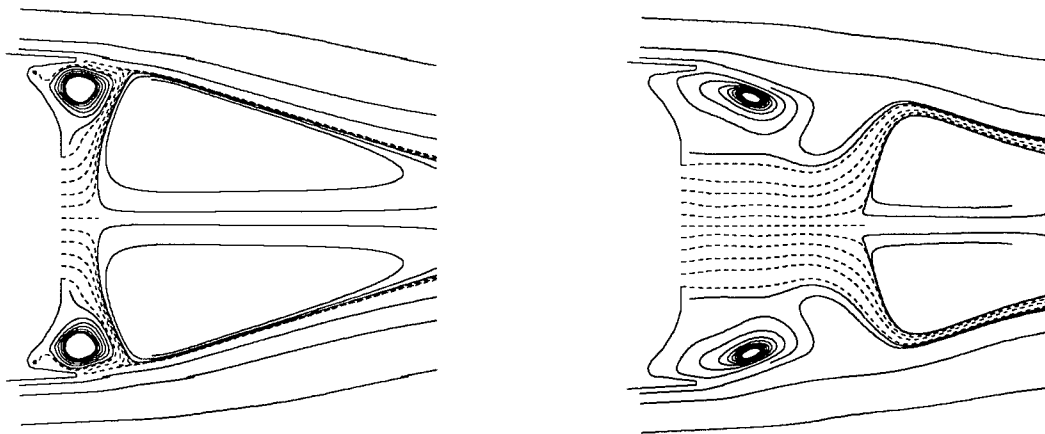
Figure 10: Partical Traces, M=2.0, T_{o_j}=300K   Figure 11: Partical Traces, M=2.0, T_{o_j}=1200K

base region for bleed gas temperatures of 300K and 1200K, respectively. The dashed line traces are particles which originate from the bleed ejection hole. The lighter traces were released in other parts of the flow field. The effect of temperature is rather dramatic in changing the recirculation pattern within the base flow field. At the lower temperature, the bleed gas is shown to be captured by the recirculation region near the base. As the temperature is increased, all of the bleed gas moves downstream. The higher temperature gives rise to an increased velocity and thus an effective increase in the mass injection parameter.

The results of these computations, in terms of the linear effect of injection, are shown in Figure 12. The data are unique in that they illustrate the combined effect of Mach number and injected gas temperature. The computed data at an injection temperature of 1500 K are consistent with the only data available.[14] At all stagnation temperatures of the gas, the slope $[\frac{d(P_b/P_\infty)}{dI})_{I=0}$ increases rapidly with Mach number and almost linearly with temperature. An approximate formula which describes the data is:

Figure 12: $\frac{d[P_b/P_\infty]}{dI}$ vs M_∞, $T_{o_j} \geq 1200K$

$$[\frac{d(P_b/P_\infty)}{dI}]_{I=0} = (-5.3953 + 0.01723T_j)M_\infty$$
$$+(4.6101 - 0.01463T_j)M_\infty^2$$
$$+(-0.5660 + 0.00446T_j)M_\infty^3. \quad (8)$$

Note that with the linear model (Equation 5) of the effect of injection on base pressure, it is possible to correct the no injection base drag to account for base bleed. At each step in a trajectory calculation where an estimate of Mach number, free-stream static pressure, spin rate and inert (no injection) drag coefficients are specified, then Equations (1), (2), and (4), can be solved simultaneously to find the change in base pressure (and thus base drag) with injection.

IV. RESULTS

The theory developed in the previous sections has been used to develop a computer program which includes a two dimensional, modified point mass trajectory model based on the work of Lieske et. al.[15] The resultant technique has been compared to several flight test cases. The primary case is that of an instrumented[16] M864 in which the base and combustion chamber pressures were measured in flight. Radar data also have been obtained by Lieske[17] over a range of launch conditions. These results are compared to the present model's predictions.

1. Instrumented Flight Test Case

Kayser, et al[16] launched an instrumented M864 in September 1988. The projectile contained pressure transducers to monitor the static pressure at two points on the base and the stagnation pressure in the combustion chamber. In addition, a thermocouple measured the temperature in the combustion chamber. These data were telemetered to recording stations on the ground. The projectile was not a standard M864 in that the mass was 81.9 per cent of normal. The launch conditions were Mach 1.30 at a quadrant elevation of 850 mills.

Figure 13 shows the main result of the experiment which is the pressure-time histories for two points on the base and in the chamber. The origin of the time scale is with respect to the exit of the shell from the gun. The initial base pressure corresponds to the inert projectile at Mach number of 1.30 . The two base pressure taps indicated almost the same pressure throughout the burning phase, although there was a significant difference after burnout. A significant pressure jump was observed as the projectile decelerates through Mach number one at 8. to 9. seconds into the flight. The chamber pressure measurements roughly parallel the base pressures but at a some what higher level as needed to produce the gas generator exhaust.
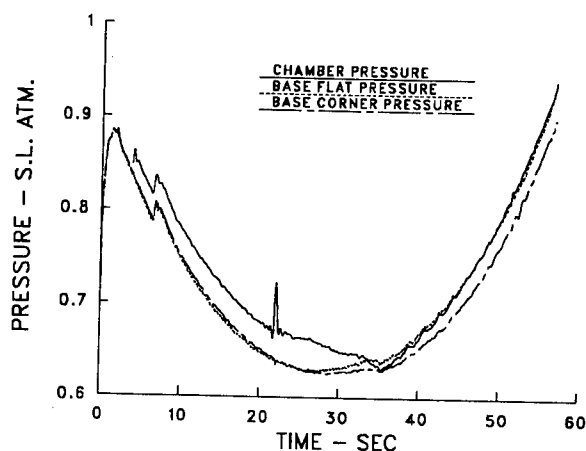
Figure 13: Pressures Measured[16] in Flight
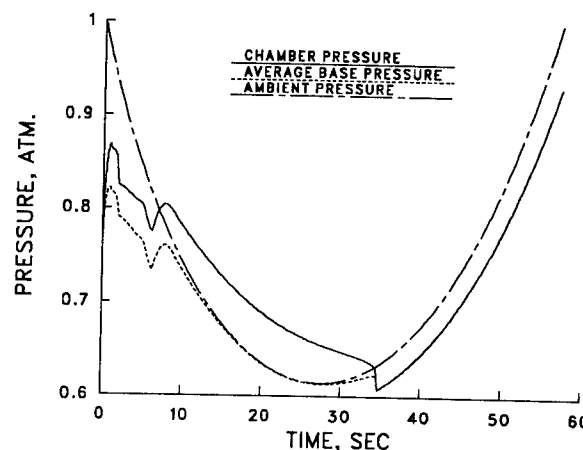
Figure 14: Computed Pressures

The next figure, 14, shows the computed pressure histories using the present analysis. The results are in resonably good accord with the previous experimental data. There are some discrepancies including that the base pressure did not jump up as high initially. This would seem to indicate that the initial drag reduction is not as great as obtained in flight. A third line on this figure indicates the local ambient static pressure along the trajectory. The base pressure actually becomes larger than the local static

pressure indicating a thrust is being generated near the apogee. The chamber pressure history also roughly parallels the base pressure. The burnout is clearly observed at 36.8 seconds.

Table 1: PREDICTED FLIGHT RANGE DATA FOR M864

QE=850. MILS				
CHARGE	M_i	RANGE COMPUTED	RANGE FIRING TABLES	DIFFERENCE
		km	km	PER CENT
6W-	1.30	12.82	12.557	+2.0
7W	1.61	17.87	17.153	+4.2
7R	1.97	22.24	21.996	+1.1
8R	2.37	27.32	28.129	-2.9

Figure 15 shows a comparison between the predicted trajectory for this round with the motor on and with motor off. The effect of drag reduction caused by the gas generator results in a 23 per cent increase in range.

2. Comparison with Firing Table Data

In order to develop firing information for the M864, BRL has been conducting flight tests of production ammunition over a wide range of conditions. The flight experiments employ Doppler radar to track the trajectory. From these measurements and 3D modified point mass simulation, the drag and velocity of the projectile in flight are deduced. The data reduction technique and results have been reported by Lieske.[17] The modeling analysis developed here has been applied to some of these cases. A picture of the ability of the model to predict M864 flight performance can be obtained from Table 1 in which the range as determined by Firing tables is compared with the current model. The table also includes a fourth case of the light weight projectile launched with a non-standard charge. This table

Figure 15: Predicted M864-L trajectories

shows that the computation gives the correct trend and magnitude. The present model

agrees with that predicted by Firing Tables computation within 800 m for the four cases considered.

V. CONCLUDING SUMMARY

An engineering model for the performance of the M864 base burn projectile has been assembled and tested on a number of flight cases. The model takes into account the three main elements of the problem: gas generator mass flow prediction, the drag reduction caused by the mass injected into the wake, and a two dimensional, modified point mass trajectory computation. This new predictive capability provides qualitative predictions of all the physical aspects of the system's performance. The computer model is based on laboratory experiments, analysis, and flight tests of the inert shell. No special factors have been introduced to improve the agreement with measured data from flights of the M864 projectile with base bleed.

The Navier-Stokes computations of the base flow have been a critical element in formulating a comprehensive model. These computations have been used to provide a correlation of change in base pressure with mass injection, Mach number and taking into account the temperature of the gas. This later effect is of critical importance for the development of a physically realistic model of the effect of mass addition on drag reduction. Applicable data have been unavailable in previous experimental studies.

The analytical model of the M864 has been compared with an instrumented flight test case. The measured chamber pressure and base pressure versus time are predicted by the model. The computed burnout time predicted was 36.8 seconds compared to an observed burnout of 35.3 seconds. The time of flight is also predicted within one second. The model was also applied to a number of flight conditions representing normal operational situations. Doppler radar data are available for these conditions and total range predictions gave correct trends and range values within 4.2 per cent of the test results.

The developed engineering model provides a significant new tool for the analysis of future designs. Current programs in extended range artillery are using this analysis for initial design guidance. The impact in financial savings using this methodology can be great given that parametric design changes can now be studied before hardware is produced.

References

1. S.N.B. Murthy, (Ed.), Aerodynamics of Base Combustion, Progress in Astronautics and Aeronautics, Vol. 40, AIAA, New York, New York, 1976.

2. K.K. Kuo, J.N. Fleming, "Proceedings of First International Symposium on Special Topics in Chemical Propulsion: Base Bleed," Athens, Greece, November, 23-25, 1988.

3. M.S. Miller and H.E. Holmes, "An Experimental Determination of Subatmospheric Burning Rates and Critical Diameters for AP/HTPB Propellant," Proceedings of the 1987 JANNAF Combustion Meeting, Monterrey, California, October 1987.

4. L.D. Kayser, J.D. Kuzan, and D.N. Vazquez, "Ground Testing for Base-Burn Projectile Systems," BRL-MR-3708, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, November 1988 (AD A201107).

5. L.D. Kayser, "Effects of Base Bleed and Supersonic Nozzle Injection on Base Pressure," Memorandum Report No. 2456, USA Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland, March 1975 (AD B003442L).

6. C.J. Nietubicz, and J. Sahu, "Navier-Stokes Computations of Base Bleed Projectiles," Paper No. II-2, First International Symposium on Special Topics in Chemical Propulsion: Base Bleed, Athens, Greece, November, 23-25, 1988.

7. J. Sahu, C.J. Nietubicz, and J.L. Steger, "Navier-Stokes Computations of Projectile Base Flow with and without Mass Injection," AIAA Journal Vol. 23, No. 9, September 1985, pp. 1348-1355.

8. Nietubicz, C. J., Pulliam, T. H., and Steger, J. L., "Numerical Solution of the Azimuthal-Invariant Navier-Stokes Equations," U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, ARBRL-TR-02227, March 1980. (AD A085716)

9. Beam, R., and Warming, R. F., "An Implicit Factored Scheme for the Compressible Navier-Stokes Equations," AIAA Paper No. 77-645, June 1977.

10. Sahu, J., and Nietubicz, C. J., "Improved Numerical Prediction of Transonic Flow," Proceedings of the 4th Army Conference on Applied Mathematics and Computing, Cornell University, May 1986.

11. Sahu, J., Nietubicz, C. J., and Steger, J. L., "Navier-Stokes Computations of Projectiles Base Flow with and without Base Injection," U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, ARBRL-TR-02532, November 1983. (AD A135738)

12. Baldwin, B. S., and Lomax, H., "Thin-Layer Approximation and Algebraic Model for Separated Turbulent Flows," AIAA Paper No. 78- 257, 1978.

13. Nietubicz, C. J., Heavey, K. R., and Steger, J. L., "Grid Generation Techniques for Projectile Configurations," Proceedings of the 1982 Army Numerical Analysis and Computers Conference, ARO Report 82-3, February 1982.

14. Z. Ding, Y. Liu and S. Chen, "A Study of Drag Reduction by Base Bleed at Subsonic Speeds," First International Symposium on Special Topics in Chemical Propulsion: Base Bleed, Athens, Greece, November, 23-25, 1988.

15. R.F. Lieske, and M.L. Reiter, "Equations of Motion for a Modified Point Mass Trajectory," BRL Report No. 1314, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, March 1966. (AD 485869)

16. L.D. Kayser, J.D. Kuzan, and D.N. Vazquez, "Flight Testing for 155MM Base Burn Projectile," BRL-MR-3708, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, November 1988. (AD A201107)

17. R.F. Lieske, "Determination of Aerodynamic Drag and Exterior Ballistic Trajectory Simulation for the 155mm, DPICM, M864 Base-Burn Projectile," BRL-MR-3768, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, June 1989. (AD A209510)

Elastic Constants of a Graphite-Epoxy Composite by Ultrasonic Wave Velocity
Measurements (U)

Dattatraya P. Dandekar
Anthony G. Martin
U.S. Army Materials Technology Laboratory
Watertown, MA. 02172-0001

I. INTRODUCTION

Continuous fiber reinforced polymeric composites are often designed to satisfy the specific loading requirements of a structure . The adopted design is usually based on methods to determine elastic properties of the composite from the elastic properties and geometrical layups of its constituents [1,2]. These methods have provided useful design data and have pointed out difficulties associated with such an endeavor especially when information about the elastic properties of a thick composite are to be obtained. Two sources, relevant to these difficulties, are non-uniformity of layups used in the fabrication of such a composite and paucity of elastic wave velocity data as a function of wave frequency. Both of these sources contribute to inappropriate prediction of the elastic constants of the composite and thereby its performance under static or dynamic loading conditions. The objectives of the present work are to determine the variability in the measured values of elastic constants of thick graphite epoxy composite specimens obtained from a bar and to determine the variation in its measured values of elastic constants due to frequencies of sound wave, in other words to measure dispersion of sound waves in the composite.

II. MATERIAL

The graphite-epoxy composite investigated in the present work was fabricated from Hercules IM7-8551-prepreg material. The IM7 fiber bundle diameter was around 5μm. The thickness of the cured prepreg was 0.127$\pm$0.013μm. The prepregs were arranged in $[0^{O}/\pm45^{O}/0^{O}]_{s}$. s being equal to 36. The 0° orientation of fibers was designated to be along the x-axis and prepregs were laid in the x-y plane. The stacking direction of the prepregs was along the z-axis, (Fig. 1). The approximate dimensions of a rectangular bar of the graphite-epoxy composite were 32 cm x 6.3 cm x 2.1 cm. An examination of the layup in this composite suggested it to be orthotropic. This formed the basis for conducting measurements of sound wave velocities in various specific directions of this composite. The assumption of orthotropic symmetry implied that the elastic property of the composite can be described

GRAPHITE - EPOXY STRUCTURE
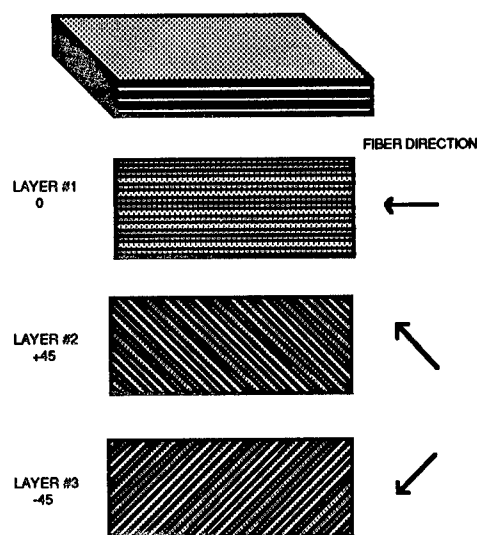
GRAPHITE - EPOXY SPECIMENS
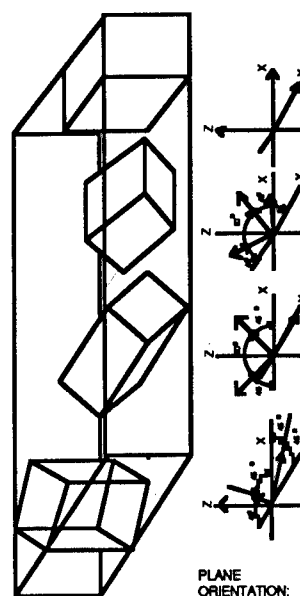
Figure 1(a) A schematic of graphite-epoxy prepreg layups in the composite.
(b) Orientation of various specimens obtained from a bar of the composite.

completely by nine independent elastic constants. In other words, at least nine independent elastic wave velocity measurements are required to determine the values of these constants of the composite. The density of the specimens of the composite varied between 1.491 ± 0.005 and 1.564 ± 0.005 Mg/m^3. Its average density was $1.53Mg/m^3$.

III. EXPERIMENTAL PROCEDURE

For elastic constant measurements, phase velocities of ultrasonic waves are required. Phase velocity is defined as the velocity of individual cycles in a continuous wave, and is given as

$$V = f\lambda = \omega/k \qquad (1)$$

where V is the phase velocity, f is the frequency of sound wave, λ is the wavelength, ω is the angular frequency, i.e., $2\pi/\lambda$. If the phase velocity is non-dispersive, i.e., it does not vary with frequency in a material, then its elastic constants remain unchanged under static or dynamic loading conditions for infinitesimally small strains.

The velocity of sound waves is affected by the dimensional relationship between the wavelength and the specimen. At low frequency the wavelength is large. If the wavelength is much larger than the cross-section dimensions of the specimen then the phase velocity at low frequency is (V_L) is given by

$$V_L = (\rho S)^{-1/2} \qquad (2)$$

where ρ is the density and S is the appropriate elastic compliance.

The requirement for low frequency measurements is that the wavelength of the ultrasonic wave be 5 or more times larger than the diameter (D) of the specimen.

At a high frequency, with wavelength being much smaller than the cross-sectional dimension, the phase velocity (V_H) is given by

$$V_H = (C/\rho)^{1/2} \qquad (3)$$

where C is the appropriate elastic constant.

This relation is based on the propagation of a plane wave front. The wave fronts are considered to be plane when the dimension of the specimen in the wave propagation direction, is within the Fresnel region (F), i.e.,

$$d \leq F = r^2/\lambda \qquad (4)$$

where r is the radius of the transducer generating plane wave excitation.

Table 1 summarizes the requirements on the values of the effective diameter and the maximum thickness of a specimen needed to carry out phase velocity measurements at various frequencies.

Table 1   Requirements on the values of the effective diameter (D) and thickness (F) of a specimen to carry out wave velocity measurements at various frequencies.

Velocity (km/s)	Frequency (MHz)								
	0.1			1.0			2.0		
	λ (mm)	D (mm)	F (mm)	λ (mm)	D (mm)	F (mm)	λ (mm)	D (mm)	F (mm)
1	10	2	10	1	0.2	100	.5	0.1	200
5	50	10	2	5	1.0	20	2.5	0.5	40
10	100	20	1.0	10	2.0	10	5.0	1.0	20

+The values of F are calculated with the radius of the transducer being 10 mm.

Phase measurements were made at frequencies from about 100 kHz to about 2 MHz.

For these phase measurements, an image superposition method[3] similar to the pulse - echo overlap method[4] was used. This method employs bursts of ultrasonic vibrations rather than continuous waves . The bursts consist of a continuous wave amplitude-modulated by sinusoidal pulses synchronized with the wave. The repetition rate of the pulses is 1/2048 times the frequency of the continuous wave. Along with the sinusoidal envelope of the pulses, their duration is made long enough to encompass many cycles of the wave in order to make it as monochromatic as possible. Images of the pulses are superposed by control of the timing of pulses relative to the timing of oscilloscope sweeps. The control of timing is done by means of digital circuitry.

Phase data are obtained by comparing the phase of individual cycles as they enter and as they leave a specimen. This is done by adjusting the advance of bursts applied to the specimen so that the lead time equals the travel time of individual cycles. Then the images of individual cycles coincide as they are presented alternately on the oscilloscope.

These measurements are done with two pairs of transducers as nearly identical as possible. One pair is coupled together. The other pair is separated by and coupled to the specimen. One transducer of each pair is connected in parallel to the signal source. The other two transducers are connected to the two signal inputs of the oscilloscope through two preamplifiers.

The signal frequency is started at the lowest frequency at which the travel time equals one half-cycle or one cycle. From that point, the frequency is gradually increased and recorded along with the number of cycles of delay needed to keep the images superposed. An additional record may be kept of the number of cycles needed to match the envelope of the burst for group velocity determination.

For each set of data, n/L (number of cycles/specimen length i.e. $1/\lambda$) was plotted vs. frequency. In some sets of data, a linear function provided a good least squares fit. Other sets required quadratic or cubic functions of n/L. In all cases, phase data were adjusted so as to bring the intercept to the origin. This procedure is justified since if frequency is lowered, wavelength becomes very large and wave number, k, and number of cycles, n, approach zero. At these low frequencies , phase velocity is affected by the specimen cross-section size and the data are not reliable. The details of this technique are given in reference[3].

The sound wave speeds at 100 kHz were measured by means of a magnetostrictive wire transducer, in which one end is inserted into an excitation coil and the other end is bonded to a long , thin specimen. The coil is connected in parallel with the signal generator and the two preamplifiers. The echoes from within the specimen are matched with a reversal of one image, since the impedance of the specimen is less than the impedance of the wire. This type of transducer is generally used to measure longitudinal wave speeds in long , thin specimens. In the present work it has been used to measure shear wave speeds by bonding the transducer to a thin specimen normal to the axis of the thin specimen.

Table 2: Wave velocity types and their relations to elastic constants of an orthotropic composite as expressed by the products of density (ρ) and squared value of wave velocity (V).

	Wave Velocities		$\rho V^2 =$	Equation
Mode	Direction Propagation	Particle Motion		
L	$\langle 100 \rangle$	$\langle 100 \rangle$ V_1	C_{11}	6
S	$\langle 100 \rangle$	$\langle 010 \rangle$ V_2	C_{66}	7
S	$\langle 100 \rangle$	$\langle 001 \rangle$ V_3	C_{55}	8
L	$\langle 010 \rangle$	$\langle 010 \rangle$ V_4	C_{22}	9
S	$\langle 010 \rangle$	$\langle 100 \rangle$ V_5	C_{66}	10
S	$\langle 010 \rangle$	$\langle 001 \rangle$ V_6	C_{44}	11
L	$\langle 001 \rangle$	$\langle 001 \rangle$ V_7	C_{33}	12
S	$\langle 001 \rangle$	$\langle 100 \rangle$ V_8	C_{55}	13
S	$\langle 001 \rangle$	$\langle 010 \rangle$ V_9	C_{44}	14
QL	$\langle 110 \rangle$	$\langle 110 \rangle$ V_{10}	$0.5\, C_{66} + 0.25\,(C_{11} + C_{22}) + 0.5\,[\,(C_{12} + C_{66})^2 + 0.25\,(C_{22} - C_{11})^2\,]^{1/2}$	15
QS	$\langle 110 \rangle$	$\langle \bar{1}10 \rangle$ V_{11}	$0.5\, C_{66} + 0.25\,(C_{11} + C_{22}) - 0.5\,[\,(C_{12} + C_{66})^2 + 0.25\,(C_{22} - C_{11})^2\,]^{1/2}$	16
S	$\langle 110 \rangle$	$\langle 001 \rangle$ V_{12}	$0.5\,(C_{55} + C_{44})$	17
QL	$\langle 101 \rangle$	$\langle 101 \rangle$ V_{13}	$0.5\, C_{55} + 0.25\,(C_{11} + C_{33}) + 0.5\,[\,(C_{13} + C_{55})^2 + 0.25\,(C_{11} - C_{33})^2\,]^{1/2}$	18
QS	$\langle 101 \rangle$	$\langle \bar{1}01 \rangle$ V_{14}	$0.5\, C_{55} + 0.25\,(C11 + C_{33}) - 0.5\,[\,(C_{13} + C_{55})^2 + 0.25\,(C_{11} - C_{33})^2\,]^{1/2}$	19
S	$\langle 101 \rangle$	$\langle 010 \rangle$ V_{15}	$0.5\,(C_{66} + C_{44})$	20
QL	$\langle 011 \rangle$	$\langle 011 \rangle$ V_{16}	$0.5\, C_{44} + 0.25\,(C_{22} + C_{33}) + 0.5\,[\,(C_{23} + C_{44})^2 + 0.25\,(C_{22} - C_{33})^2\,]^{1/2}$	21
QS	$\langle 011 \rangle$	$\langle 0\bar{1}1 \rangle$ V_{17}	$0.5\, C_{44} + 0.25\,(C_{22} + C_{33}) - 0.5\,[\,(C_{23} + C_{44})^2 + 0.25\,(C_{22} - C_{33})^2\,]^{1/2}$	22
S	$\langle 011 \rangle$	$\langle 100 \rangle$ V_{18}	$0.5\,(C_{66} + C_{55})$	23

431

Using 1,2, and 3 to represent the x,y,and z axes,the elastic constants matrix [C] for an orthotropic material is

$$[C] = \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{22} & C_{33} & 0 & 0 & 0 \\ C_{13} & C_{23} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{bmatrix} \quad (5)$$

The elastic compliance matrix [S] is inverse of matrix [C].

The ultrasonic wave velocities were measured in <100>, <010>, <001>, <110>, <101>, and <011> directions to determine the nine independent elastic constants of the graphite epoxy composite. The nature of these velocities and their relations to the elastic constants are given in Table 2. These relations are obtained from the Christoffel relation for wave propagation in an orthorhombic i.e. orthotropic solid[5]. In this table L and S denote longitudinal and shear waves, respectively and prefix Q to them denote the quasi nature of these respective waves. The frequencies at which these velocities were measured varied between 0.1 to 2 MHz. Wave velocity measurements could not be made at higher frequencies due to attenuation of waves in the composite specimen. The precisions of the longitudinal and shear wave velocities measurements, based on repeated measurements of the velocities on a single specimen,are estimated to be 1 and 1.5 respectively.

A wire transducer was used to measure longitudinal and shear wave speeds only in <100> and <010> directions due to limitation of the bar dimensions of the graphite epoxy composite. The lengths of these specimens were 9.5 cm and 6 cm, respectively. Their cross-sections were 0.2 x 0.2 cm.

IV RESULTS AND DISCUSSION

The results of wave velocity measurements are presented

(i) to show the extent of variability in the measurement of wave velocities in different specimens of the graphite epoxy composite,

(ii) to show the dispersion of ultrasonic waves with frequency of the wave in the composite,

(iii) to determine the values of the nine independent elastic constants from these wave velocity measurements at various frequencies,and

(iv) to compare the results of higher ultrasonic measurements with those obtained at lower frequencies and static measurements. The measured values of wave velocities in eight different specimens of graphite epoxy composite are given in Table 3. As indicated in Table 2, longitudinal and shear modes of propagation along <100>, <010>, and <001> are pure i.e., the wave propagation vector is the same as the orientation. However, even for these modes, the variations in the measured velocities exceed the estimated errors in these measurements from specimen to specimen. For these orientations, it is clear from Table 2 that the following relations must hold.

$$V_2 = V_5 \qquad\qquad (24)$$
$$V_3 = V_8 \qquad\qquad (25)$$
$$V_6 = V_9 \qquad\qquad (26)$$

Table 3   Measured wave velocities in eight specimens of graphite epoxy composite at 2MHz

Specimen	Orientation	Thickness (mm)	Velocities (km/s) Longitudinal	Shear (1)	Shear (2)
1	<100>	25.27	7.210±0.108	3.727±0.035	1.656±0.007
	<010>	20.24	3.428±0.086	3.739±0.118	1.445±0.005
	<001>	20.22	2.497±0.004	1.597±0.004	1.367±0.006
2	<100>	25.62	8.921±0.134	3.541±0.061	1.746±0.006
	<010>	20.21	3.619±0.054	3.541±0.092	1.482±0.006
	<001>	20.21	2.691±0.004	1.633±0.011	1.449±0.020
3	<100>	25.39	7.337±0.146	3.593±0.052	1.651±0.013
	<010>	20.13	3.069±0.086	3.565±0.056	1.442±0.007
	<001>	20.29	2.494±0.006	1.591±0.006	1.332±0.004
4	<100>	13.05	7.323±0.146	--	--
5	<010>	12.48	3.709±0.074	--	--
8	<110>	26.93[+]	6.301±0.085	3.181±0.047	1.653±0.047
	<110>	26.15[+]	5.932±0.074	2.954±0.025	1.577±0.023
7	<101>	13.89[+]	5.5651±0.078	2.185±0.040	2.690±0.046
	<101>	13.70[+]	5.4263±0.099	2.147±0.043	2.757±0.107
6	<011>	13.68[+]	3.092±0.046	1.896±0.022	2.893±0.139
	<011>	13.63[+]	3.111±0.046	1.924±0.014	3.106±0.253

+   Lateral dimensions of these specimens varied between 18.86 and 19.83 mm.

These, in general, appear to hold though certainly not as well as in single crystal specimens. Similar variability is also observed in <110> & <110>,<101> &<101>, and <011> &<011>paired orientations.In other/words, whereas measurements of these velocities are in general of similar magnitudes, the variability in these measurements exceeds the precision of each measurement and the variation in the density of the specimen, assuming elastic constants remain unchanged.

The ultrasonic measurements indicated that all the shear wave propagation velocities in the graphite epoxy composite are dispersionless. Of the longitudinal waves, all except the one propagating normal to the layup direction i.e. in z direction are dispersive. The dispersive nature of these longitudinal modes of propagation namely V_1, V_4, V_{10}, V_{13} & V_{16} can be represented by the relation.

$$f = a\lambda^{-1} + b\lambda^{-2} + c\lambda^{-3} \qquad (27)$$

where f is frequency of the wave in MHz, λ is in cm., and a, b , and c are constants.

Table 4   Constants for the dispersive longitudinal waves in graphite epoxy composite.

Specimen	Propagating Direction	a	b	c
1	<100>	8.188×10^5	-3.527×10^4	0
3	<100>	8.944	-5.896	0
4	<100>	9.350	-7.420	0
1	<101>	4.870	-3.090	1.062×10^3
2	<010>	5.085	-3.267	1.111
3	<010>	4.967	-2.798	0.356
5	<010>	4.938	-2.679	0.741
6	<011>	3.864	-1.170	0
	<011>	3.734	-0.976	0
7	<101>	6.070	-1.405	0
	<101>	5.887	-1.250	0
8	<110>	7.366	-3.355	0
	<110>	6.927	-2.952	0

The constant a is the expected phase velocity of these longitudinal modes at zero frequency. The values of these three coefficients obtained from the measured values of f and wave number for these modes in various specimens of graphite epoxy composite are given in Table 4. This table shows that whereas the coefficients b for these modes in different specimens are of similar order of magnitudes the expected zero frequency velocities could be quite different from specimen to specimen. This is most strikingly shown by the longitudinal wave velocity in <100> direction. A general feature of these dispersion curves is that phase velocity of these modes decreases with an increase in frequency to 2 MHz.

Table 5   Average wave velocities in graphite epoxy composite at various frequencies.

| Wave Type | Wave Velocity (km/s) Frequency (MHz) | | | |
	0	0.1	1	2
V_1	8.188±0.122	8.145	7.732	7.210
V_2	3.620±0.100			
V_3	1.684±0.050			
V_4	5.005±0.141	4.844	4.368	3.648
V_5	3.615±0.11			
V_6	1.456±0.032			
V_7	2.561±0.061			
V_8	1.608±0.023			
V_9	1.383±0.060			
V_{10}	6.814±0.080	6.783	6.493	6.134
V_{11}	3.068±0.036			
V_{12}	1.615±0.035			
V_{13}	6.058±0.088	6.031	5.779	5.468
V_{14}	2.166±0.042			
V_{15}	2.724±0.076			
V_{16}	3.793±0.046	3.816	3.486	3.101
V_{17}	1.910±0.018			
V_{18}	3.000±0.196			

In order to estimate the values of the elastic constants of the composite from the wave velocity measurements made on the various specimens it was necessary to consolidate the data presented in Tables 3 and 4 which would be representative of the material. Shear waves and the single longitudinal wave V_7 did not present any problem, since they were dispersionless,

435

Table 6   Internal consistencies of wave velocities in the graphite epoxy composite for both dispersive and nondispersive modes.

Relation	Frequency (MHz)	Left Hand Side	Right Hand Side
Nondispersive modes			
24		3.62±0.10	3.615±0.11
25		1.68±0.05	1.61±0.02
26		1.46±0.03	1.38±0.06
28		2.61±0.11	2.36±0.11
29		7.42±0.41	7.55±0.11
30		9.00±1.18	7.90±0.09
Dispersive modes			
31	0.0	55.84±2.4	59.13±3.16
	0.1	55.42±2.4	57.99±3.1
	1.0	51.57±2.2	52.52±2.7
	2.0	47.04±2.0	45.73±2.27
32	0.0	41.39±0.97	39.51±2.20
	0.1	41.06±0.97	39.16±2.17
	1.0	38.09±0.97	35.88±1.92
	2.0	34.59±0.97	31.98±1.60
33	0.0	18.03±0.76	17.82±0.87
	0.1	17.82±0.76	17.52±0.86
	1.0	15.80±0.67	14.83±0.73
	2.0	13.26±0.56	11.95±0.59

and the representative values of these velocities were calculated to be as their respective arithmatic average. For the remaining waves showing dispersion, a best fit through the individual curves was made to represent its respective modes. All the values of the various wave velocities thus obtained are tabulated as the average velocities in Table 5. The subsequent task was to examine if these velocities satisfy the compatible relations pertaining to assumed orthotropic symmetry for the composite. There are nine such relations; three of them (24)-(26) have been already mentioned above. The remaining relations are given below.

$$V^2_{12} = 0.5 \ (V^2_3 + V^2_6) \qquad (28)$$
$$V^2_{15} = 0.5 \ (V^2_2 + V^2_6) \qquad (29)$$

$$V^2{}_{18} = 0.5 \, (V^2{}_2 + V^2{}_3) \qquad (30)$$
$$V^2{}_{10} + V^2{}_{11} = V^2{}_2 + 0.5 \, (V^2{}_1 + V^2{}_4) \qquad (31)$$
$$V^2{}_{13} + V^2{}_{14} = V^2{}_3 + 0.5 \, (V^2{}_1 + V^2{}_7) \qquad (32)$$
$$V^2{}_{16} + V^2{}_{17} = V^2{}_6 + 0.5 \, (V^2{}_4 + V^2{}_7) \qquad (33)$$

A tabulation of these calculations is given in Table 6. Table 6 shows that these relations are satisfied, except for possibly the relation (30). A further ground for holding that the material has orthotropic symmetry is the fact the independent measurements of the velocity modes showing dispersion satisfy the orthotropic symmetry requirement as well as shown by the relations (31)-(33). Thus the assumption of orthotropic symmetry for the composite is validated.

Calculations of the values of the nine independent elastic constants from the 18 wave velocity values given in Table 5 were based on the following considerations. First the values of the constants must be representative of pertinent velocity measurements. Second the effect of inordinately large discrepancies showed up by relation (30) on the values of the elastic constants should be minimized. These two considerations led to the following steps for the calculations of various elastic constants.

(i) Since the shear wave velocities were dispersionless, the values of C_{44}, C_{55}, and C_{66} were determined by the method of least squares using the values of V_2, V_3, V_5, V_6, V_8, V_9, V_{12} and V_{15} as described in reference 6.

Table 7   Values of adiabatic elastic constants (C_{ij}) of graphite epoxy in units of GPa.

Elastic Constants	Frequency (MHz)			
	0	0.1	1.0	2.0
C_{11}	102.58	101.50	91.47	79.54
C_{22}	38.33	37.40	29.19	20.36
C_{33}	10.03	10.03	10.03	10.03
C_{44}	3.12	3.12	3.12	3.12
C_{55}	4.23	4.23	4.23	4.23
C_{66}	19.97	19.97	19.97	19.97
C_{12}	26.67	24.80	19.28	11.46
C_{13}	11.87	11.84	12.24	12.60
C_{23}	5.24	5.61	5.69	4.41

(ii) Since the longitudinal wave velocities in <100>, <010> and <001> are pure, their values were used to calculate estimates of C_{11}, C_{22}, and C_{33} respectively.

(iii) The values of off-diagonal constants C_{12}, C_{13} & C_{23} were determined from the values of quasi-longitudinal and quasi-shear wave velocity pairs (V_{10}, V_{11}), (V_{13}, V_{14}), and (V_{16}, V_{17}), respectively.

The values of elastic constants determined in the manner described above are given in Table 7. The errors associated with the calculated values of C_{11}, C_{44}, C_{55}, and C_{66} are 3 per cent and the errors associated with the values of the remaining constants are about 4.5 per cent. This table shows that whereas the values of C_{33}, C_{44}, C_{55}, and C_{66} remain unchanged at various frequencies, the values of C_{11}, C_{22}, C_{12}, C_{13}, and C_{23} change with frequency. The value of density of the graphite-epoxy used in the above calculations is 1.53 Mg/m^3.

Finally, how valid are these estimates of elastic constants especially at low frequencies i.e., 0 and 0.1 MHz? To test the validity of these constants five static compression experiments and four wave velocity measurements using a wire transducer were performed. Static compression experiments were performed along the <100> direction by E. Goeke[7] and along the <011> and <101> directions of the composite by R. Swanson[8]. The static experiments were performed on specimens 1- 3 , 6, and 7 used in ultrasonic experiments (Table 3). The static experiments thus eliminated the possibility of variability in the values of elastic constants due to nonhomogeneity of the composite bar as evidenced by the measured density variation mentioned earlier (Section II). On the other hand , flexural and shear wave velocity measurements at 0.1 MHz using the wire transducer could only be performed in long thin specimens of the composite along the <100> and <010> directions because of geometrical limitations of the graphite-epoxy bar. These measurements yielded the values of Young's (E) and shear (C) moduli along these directions. The values of these constants from these two types of experiments along with their respective estimates calculated from ultrasonic wave

Table 8   Values of Young's (E) and shear (C) moduli (in units of GPa) and Poisson's ratio at 0.1 and 0 MHz from ultrasonic and wire transducer methods & static compression measurements.

| | Frequency (MHz) | | | |
| | 0 | | 0.1 | |
	Ultrasonic	Static	Ultrasonic	Wire
E <100>	77±7	87±7	78±7	74.2±3
E <010>			30.5±2.7	19.3±0.7
E <011>	8.8±0.8	10.2±0.5		
E <101>	11.6±1.2	10.0±0.5		
C44			3.12±0.09	3.04±0.09
C_{55} (GPa)			4.22±0.13	4.26±0.1
v_{12}	0.231	0.12-0.247		

velocity measurements are given in Table 8. This table shows that all elastic parameters except for Young's modulus along <010>, i.e., E<010> determined from measurements at low (0.1 MHz) and at high (> 0.1 MHz) ultrasonic frequencies are within the precisions of their respective measurements. These results provide a severe test of the validity of the procedure adopted to predict the values of elastic constants at low frequencies and under quasi-static loading conditions from the ultrasonic wave velocity measurements at high frequencies. These agreements are especially encouraging when one considers the fact that the estimates of

Young's modulus and Poisson's ratio (v) require calculations of the elastic compliance matrix [S] which involve the measured values of C_{11}, C_{22}, C_{33}, C_{12}, C_{13}, and C_{23}.

V. CONCLUSIONS

The results of the present work indicate that if the graphite-epoxy bar, from which different specimens with various orientations were used to measure elastic wave velocities, is representative of the fabrication technology of such a composite, then one can expect the material to possess non-uniform elastic properties. The variations in the measured elastic wave velocities given in Table 3 in different specimens with the same orientation can arise due to (i) misorientation of fibers in prepregs, (ii) misorientations in $(0, \pm 45^{\circ}, 0)$ layups of prepregs, (iii) density variation in the bar, (iv) and specimen orientations. Since the precision of the velocity measurements by the ultrasonic pulse overlap technique amounts to 2 per cent, the observed density variation could account for another 2.4 per cent, and finally, misorientation of 0.5° amounts to 0.6 per cent error, then any observed variation in the measured velocities for a given mode exceeding 5 per cent has to be accounted for by factors (i) and (ii). However, this hypothesis needs to be confirmed by direct observation of the layups and fibers. An important implication of this conclusion is that ultrasonic techniques can be profitably used to determine, control, and assure the quality of such a composite bar in terms of its elastic properties.

A second conclusion that can be made is that the elastic constants of such a composite do vary with the frequency of ultrasonic waves and where, precise determination of the elastic response of such a material is of concern, elastic constants determined at a single frequency may lead to misleading prediction of its performance. This is amply illustrated by the values of C_{11}, C_{22}, C_{12}, C_{13}, and C_{23} at frequencies 0 to 2 MHz for this composite in Table 7. These variations in the values of elastic constants are quite large except for C_{13}. C_{13} also happens to be the only elastic constant which increases with an increase in the frequency of the ultrasonic wave. At present we do not have a satisfactory model or explanation for the observed dispersion of longitudinal waves in the x- and y-directions and the nondispersion of longitudinal waves in the z-direction and all the shear waves in this composite.

Lastly, the limited number of shear wave velocity measurements made in long thin specimens of the composite by means of the wire transducer show that it is a reliable technique for this purpose, in addition to its conventional use for the measurement of flexural wave velocity, and therefore determination of Young's modulus of a material. However, more experiments need to be performed to establish this technique on a firm basis for the measurement of shear modulus. An advantage of the technique is that it does not require a large size specimen.

REFERENCES

1. Christensen, R.M., Mechanics of Composite Materials, John Wiley (1979) and references therein
2. Tsai, S.W., Composite Design 1986, Think Composite (1986)
3. Martin, A.G., Phase Velocity Measurements in Dispersive Materials by Narrow-Band Burst Phase Comparison, Army Materials and Mechanics Research Center,Watertown,MA. AMMRC TR 76-22 (1976).
4. Papadakis, E.P., J. Acoust Soc. Am. 42 (1967), 1045
5. Love, A.E., The Mathematical Theory of Elasticity, Dover Publications (1944)
6. Chou, S.C., and Dandekar, D.P., To be published in the proceedings of International Conference on Composite Materials, ICCM -7, (1989).
7. Goeke, E., Personal Communication. Static experiments in <010> and <001> directions of these three specimens could not be performed because the compression experiments in <100> direction were continued till the specimens failed.
8. Swanson, R., Personal communication. These experiments were done at very low strains i.e., 0.4 per cent More such experiments will be performed in other directions of the composite in the future.

REDUCTION OF SWEAT ACCUMULATION RATE WITH PEDAL ANTIPERSPIRANTS (U)

Andre A. Darrigrand, MAJ,VC*, Katy Reynolds, MAJ, MC, Ronald
Jackson, CPT, MS, Murray P. Hamlet, D.V.M., Donald E. Roberts,
Ph.D
U.S. Army Research Institute of Environmental Medicine, Natick, MA
01760-5007

INTRODUCTION

Wet feet, due to accumulated perspiration, present a problem
for the soldier during operations in climatic extremes. In warm
climates, foot sweat will accumulate in the boot, moisten the
skin, and result in irritation and blisters (1). In cold
climates, moisture from sweat will accumulate in the boot,
decrease the insulation, and result in an increased incidence of
trench foot and frostbite (2). These foot injuries jeopardize
mission performance by decreasing soldier work tolerance and
increasing manpower loss, as well as increasing the burden on the
medical management system.

There are several options for reducing these sweat-induced
foot injuries. Frequent sock changes are one option. However,
this is not always practical during military operations. Another
option is to increase the air flow around the foot (3). However,
this would require a redesign of the footwear system which would
be inappropriate in cold climates. A third option is the
application of an antiperspirant to the feet, thereby reducing
sweat production and accumulated moisture in the footwear.
Antiperspirants may also reduce the coefficient of friction
between the sock and foot resulting in a decrease in the number
and severity of blisters (4).

Several case studies have suggested that antiperspirants will
reduce foot blistering in patients having congenital blister
syndromes (5, 6). Although interesting, these case reports have
not undergone the scrutiny of scientific review.

In a Canadian armed-forces study, the antiperspirant aluminum
chlorhydroxide was applied to the feet of soldiers who walked in a
cold environment (7). They reported significant reductions in
sweat accumulations in the socks after several days of
application. That investigation, however, did not report any

DARRIGRAND, REYNOLDS, JACKSON, HAMLET, ROBERTS

foot-injury data. In addition, the antiperspirant was applied via an aerosol, thereby limiting the concentration delivered to the foot (8). It might be possible to increase sweat suppression even further and for a longer duration with direct application of antiperspirant.

The purpose of this study was to evaluate the effectiveness of two antiperspirants in reducing foot sweat of soldiers walking in the heat. It was hypothesized that the antiperspirant would reduce both sweat accumulation and the coefficient of friction, thereby lessening the incidence of foot injuries.

MATERIALS AND METHODS

Subjects. A total of 19 male soldiers participated in these experiments, with 16 subjects completing each of the two phases. All of the soldiers were physically active and in excellent health with no history of skin disorders or foot problems. The subjects had a mean ($\pm$ Standard Deviation) age of 23.2 ($\pm$ 3.7) years, height of 177.1 ($\pm$ 7.1) centimeters, and weight of 76.1 ($\pm$ 8.1) kilograms. All subjects gave their voluntary and informed consent to participate in this study which was approved by the Institutional Review Board. Investigators adhered to Army Regulation 70-25 and United States Army Medical Research and Development Command Regulation 70-25 on Use of Volunteers in Research.

Protocol. Two experimental treatments were examined: application of aluminum chlorohydrate (ACH) 23.5% solution or aluminumzirconium tetrachlorohydrex glycine[1] (ZIR) 23.5%. The order of antiperspirant treatment was counter-balanced and participation in each experimental phase was separated by several months (November through February). These experiments were conducted in the late fall-winter months to minimize subject exposure to environmental heat.

Two days prior to participation in each experimental phase, the subjects legs were shaved below the knees and their legs and feet were examined for injuries. In addition, the legs and feet were examined daily during each phase and injuries were annotated. The foot/leg injuries noted were: hot spots, friction blisters, irritant dermatitis, tinea pedis, and plantar fasciitis. The clinical criteria used to diagnose these five injuries were:

    a. hot spot: A localized, superficial area of erythema present over a joint or weight bearing surface exposed to shearing forces on the skin surface.

    b. friction blister: A 1-5 millimeter circumscribed epidermal elevation containing a clear fluid usually occurring on weight bearing surfaces involving joints or areas exposed to shearing

frictional forces (9, 10).
c. <u>irritant dermatitis</u>: A dermatological disorder caused by the external application of an irritant . This disorder is manifested by diffuse erythema, swelling, and pruritus (10).
d. <u>tinea pedis</u>: A fungal infection involving the superficial layers of the skin between the toes and on the soles of the feet. The clinical presentation includes mild to moderate scaling, cracking, and pruritus.
e. <u>plantar fasciitis</u>: Inflammation and soft tissue injury involving the sole of the foot.

Each experimental phase lasted seven consecutive days with the first two days serving as control days. No antiperspirant was applied on the control days, but on the five treatment days either ACH or ZIR were applied to the feet of subjects at 0730 and 1530 hours by technicians. The subjects were not told which antiperspirant they received. The foot and leg below the knee were thoroughly moistened with a 23.5% aqueous solution of either antiperspirant and allowed to air dry. Both feet were treated to avoid side bias (11, 12); that is, to minimize the effect of one foot producing greater or lesser amounts of sweat than the contralateral foot. To standardize bathing, the subjects washed at night with a 2% suspension of Camay^R soap.
During the first (days 1 and 2) and last (days 6 and 7) two days of each study phase the subjects walked on the treadmill. The subjects walked--5.6 kilometers per hour, 0% grade--on a treadmill for one hour in a 32° centigrade, 65% relative humidity environment. For these soldiers this was a comfortable pace, and the one hour period was sufficient to elicit a profuse sweat. The three days between the control and treatment-treadmill walks allowed the antiperspirants to develop a cumulative effect. This three day period was based on underarm antiperspirant studies which have shown a maximum cumulative effect at three days (Personal communication with Mr Ike Emery of The Gillette Co., Boston, MA.). During these days, the subjects were advised to continue their normal daily activities.
Prior to entering and after leaving the environmental chamber, nude-body weights were taken to assess the subjects' hydration levels. Rectal temperature and heart rate were monitored for subject safety reasons. A flexible-rectal thermometer was inserted 10 centimeters past the anal sphincter to monitor rectal temperature. In addition, chest electrodes (CM5 placement) were positioned for monitoring heart rate and electrocardiogram. At completion of instrumentation, subjects dressed in shorts, tee shirt and hooded sweatshirt, then proceeded into the environmental chamber.
After entering the chamber, copper-constantan thermocouples

were taped to the arches of both feet to assess antiperspirant effect on foot temperature. The arch was selected because skin abrasions could occur if thermocouples were affixed to the dorsal-foot surface under the boot lace. The subjects then put on cotton socks that had been dried (at 75° to 80° C) to a constant weight. A four inch section of stockinet was worn above the sock to preclude entrance of sweat from the upper leg. Subjects wore their own combat boots because new boots often cause focal irritations until proper adjustment occurs. Before starting the treadmill walk, subjects consumed 400 milliliters of water and were encouraged to take additional water as desired. Upon exiting the treadmill, boots and socks were removed and feet examined. Subjects relaxed until heart rates and core temperatures returned to pre-walk levels. Sweat accumulation in the sock and boot was measured by comparing the pre-walk weight to that of the post-walk weight.

Statistical Analysis. Analysis of Variance with repeated measures was used to test the difference between physiological control and treatment values. Significant effects and interactions were tested by the Tukey's Test. A Paired-T Test was used to evaluate significance of skin-temperature data. Evaluation of differences of sweat accumulation between feet was made with a T test. The clinical-injury data were analyzed by the Fischer's Exact Test. Significant difference was determined at the $P<0.05$ level.

RESULTS

Sixteen subjects completed all seven days of testing for each phase. Since the desired environmental conditions were not properly controlled during the first test days (days 1 and 6) the foot sweat accumulation and skin temperature data were statistically compared only between day two (second control day) and day seven (fifth treatment day).

Figure 1 presents sweat-accumulation data for the two sessions. The total foot-sweat-accumulation was subdivided to that found in the boots and the socks. The data represent mean values for both feet as side bias was not found between left and right feet. Total sweat accumulation was decreased from control levels (mean ± SEM, 51.2 ± 4.3 g/h) by both ACH (23.2 ± 2.0 g/h) and ZIR (25.0 ± 2.7 g/h), with no difference between the two treatments. Boot sweat accumulation was decreased from control levels (26.7 ± 2.4 g/h) by both ACH (11.0 ± 1.1 g/h) and ZIR (12.2 ± 1.7 g/h) with no difference between the two treatments. Sock sweat accumulation was decreased from control levels (24.5 ± 2.4 g/h) by ACH (12.2 ± 1.2 g/h) and ZIR (12.8 ± 1.4 g/h), with no

difference between the two treatments.

Foot-skin temperature, recorded at the arch of the foot, increased from 34.6 ± 0.2^0 C before the treadmill walk to 37.7 ± 0.1^0 C after exercise. No differences for skin temperatures were found between controls and experimental conditions. Table 1 presents the number of subjects who incurred foot injuries during the control and treatment conditions. No significant differences were found between control and treatment conditions for hot spots, blisters, tinea pedis and plantar fasciitis. It should be noted, however, that overall there was a tendency for a reduced incidence of hot spots and blisters in the antiperspirant compared to control experiment. An increased incidence of irritant dermatitis was found for both ACH and ZIR compared to controls.

DISCUSSION

This study's most important finding is that both antiperspirants reduce foot sweat accumulation by approximately 50% (Figure 1). The percent reduction of foot-sweat accumulation was larger in this study than that reported (about 30% decrease) in the Canadian Armed Forces study (7). The two studies were similarly designed with the exception of the use of a cold environment by the Canadians as opposed to a warm environment in the present study. The cold environment elicited control-foot-sweat-accumulation values of approximately 48 g/h which were nearly identical to those (51.2 g/h) in the present study. Another difference between studies was the concentration of antiperspirant and the method of application. The Canadian study utilized a 4.2% solution of the drug--aluminum chlorhydroxide--in an aerosol (personal communication with Mr. Yee of Connecticut Chemical/Armstrong-Lang Laboratories, Toronto, Canada), whereas the present study employed topical application of a 23.5% aqueous solution. In addition, an aerosol antiperspirant is limited by the concentration which can clear the diameter of the spray nozzle. As a result, the greater reduction in foot sweat accumulation achieved in the present study is probably due to the direct application of a higher antiperspirant concentration.

Skin temperature values for the feet were another area of difference from the Canadian study. In that study the antiperspirant induced an insulative effect (via reduced sweating and therefore reduced heat conduction), thereby keeping the feet warmer. Their average toe temperature was increased with antiperspirant use by 1.6^0 C. (23.0^0 C. in controls versus 24.6^0 C. in treated groups). Subjects in the present study experienced an average increase of foot skin temperature of 3.1^0 C. from beginning

to end of the walk in both control and antiperspirant groups. The lack of a difference with use of an antiperspirant could be due, in part, to the contrasting climates in which the studies were conducted. Another possibility is a difference between studies in methods and techniques by which skin temperatures were acquired or even a bias in boot temperature induced by climatic differences.

This study's second most important finding was the appearance of irritant dermatitis in 38% of subjects treated with the antiperspirants, in contrast to none in the control. The irritant dermatitis reaction is most probably due to the aluminum component of the antiperspirants. Underarm antiperspirants containing aluminum salts, in concert with heat, moisture and friction, are the causative agents for much of the irritant reactions diagnosed in underarm dermatitis (13). Another consideration is the concentration of the antiperspirants. Concentrations of over-the-counter antiperspirants for underarm use are about 20% for both stick and roll-ons, while aerosols are about 8% (Personal communication with Mr. Morton of The Gillette Co., Boston, MA.). In this study, the concentration of both antiperspirants was 23.5%. Furthermore, commercial antiperspirants often contain additives which have a soothing effect upon the skin. Future research should examine the potentials of emollient additives on reducing the incidence of irritant dermatitis.

Finally, there was a tendency, although not statistically significant, for subjects to incur a lower incidence of blisters with use of both antiperspirants. It seems possible that reduced foot sweat accumulation resulted in a lowering of the coefficient of friction between the sock and skin (4). The reduced friction should decrease blister formation. In support of this idea was a trend for reduced hot spots after antiperspirant treatment.

In conclusion, this study indicates that the use of antiperspirants on the foot will: (a) reduce foot sweat accumulation and, (b) increase the incidence of irritant dermatitis. The reduced foot sweat accumulation may also help to reduce foot blistering during exercise in warm environments. In cold environments, the reduced foot sweat accumulation should be a benefit in reducing the incidence of trench foot and frost bite injuries. Future research should address changing the antiperspirant formula to reduce the incidence of irritant dermatitis and to clarify whether or not antiperspirants can reduce the occurrence of foot blisters.

REFERENCES

1. Akers, W.A. Sulzberger on Friction Blistering. Intern. J.

Dermatology, 16(5):369-372, 1977.

2. Oakley, E.H.N. The Design and Function of Military Footwear: a Review Following experiences in the South Atlantic. Ergonomics, 27(6):631-637, 1984.

3. Dewar, M.M., Crow, R.M. Sweating Characteristics of the Foot. Technical Note No. 76-11, Defence Research Establishment Ottawa, Dept. of National Defense, Canada, 1976.

4. Naylor, P.F.D. Experimental Friction Blisters. Br. J. Dermatology, 67: 327-342, 1955.

5. Tkach, J. R. Treatment of Recurrent Bullous Eruption of the Hands and Feet (Weber-Cockayne Disease) with Topical Aluminum Chloride. J. Am. Acad. Dermatology, 6 (6):1095-1096, 1982.

6. Tidman, M.J., Wells, R.S. Control of Plantar Blisters in Pachyonychia Congenita with Topical Aluminum Chloride. Br. J. Dermatology, 118 (3): 451-452, 1988.

7. Radomski, M.W., Brown, T. Effect of a Local Anhidrotic Product on Moisture Accumulation and Temperature of Feet in the Cold. DCIEM Report No. 74-R-1068, 1975, 1133 Sheppard Ave. West, P.O. Box 2000, Downsview, Ontario M3M3B9.

8. Proposed Monograph for OTC Antiperspirant Drug Products. Food and Drug Administration. Oct 10, 1978.

9. Dorland's Illustrated Medical Dictionary. W. B. Saunders Co., 27th Edition, 1988.

10. Bergfeld, W.F., Taylor, J.S. Trauma, Sports, and the Skin. Am. Journal of Industrial Medicine, 8:403-413, 1985.

11. Bromberg, A.E.C., Creutzberg, C.L., Atherton, D.J. Sweat Production in Children's Feet. Clinical and Experimental Dermatology 10: 229-232, 1985.

12. Armstrong, L.E., Matthew, W.J., Szlyk, P.C., Hubbard, R.W., Sils, I.V., Deluca, J.I., Hodenpel, J., Evans,P. Evaluation of a Field Expedient Technique for Sweat Sample Collection. Technical Report No. T2/85, USARIEM, Natick, MA., 01760-5007, NOV 1984.

13. Fisher, A.A. Contact Dermatitis ed. 3. Philadelphia, Lea & Febiger, 1986.

DARRIGRAND, REYNOLDS, JACKSON, HAMLET, ROBERTS

ENDNOTES

[1]The Gillette Company, Boston, MA 02106

DARRIGRAND, REYNOLDS, JACKSON, HAMLET, ROBERTS

TABLE 1. NUMBER OF SUBJECTS INCURRING FOOT INJURIES
WITHIN CONTROL AND ANTIPERSPIRANT EXPERIMENT

	HOT SPOTS	BLISTERS	IRRITANT DERMATITIS	TINEA PEDIS	PLANTAR FASCIITIS
C ACH	3	3	0	3	0
T ACH	2	0	7*	1	1
C ZIR	5	5	0	1	0
T ZIR	2	3	5*	2	1

* = SIGNIFICANT DIFFERENCE FROM CONTROL (P < 0.05)
C = CONTROL
T = TEST
ACH = ALUMINUM CHLOROHYDRATE
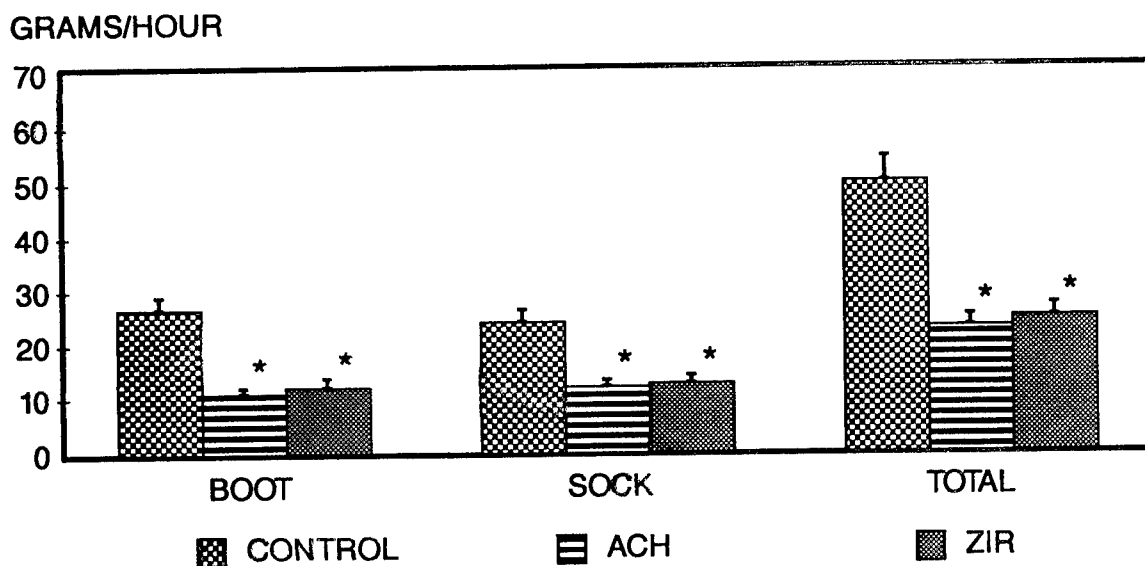ZIR = ALUMINUMZIRCONIUM TETRACHLOROHYDREX GLYCINE

DARRIGRAND, REYNOLDS, JACKSON, HAMLET, ROBERTS

**FIGURE 1. AMOUNT OF PERSPIRATION ACCUMULATED WITHIN BOOTS
AND SOCKS DURING ONE HOUR TREADMILL MARCH**

GRAMS/HOUR

ACH = ALUMINUM CHLOROHYDRATE
ZIR = ALUMINUMZIRCONIUM TETRACHLOROHYDREX GLYCINE
MEAN ± SEM   ★ = P < 0.05

INVESTIGATION OF MIR GENERATION FROM 1 MICRON LASERS

USING NONLINEAR TECHNIQUES (U)

GERALDINE DAUNT AND MICHAEL J. FERRY[*]

Laser Division, C2NVEO
Fort Belvoir, VA 22060

DR. SURESH CHANDRA

Science Applications International Corporation
McLean, VA 22102

I. Introduction

The Army is developing the highly efficient diode array pumped Nd:YAG as its "work horse" solid state laser. The Nd:YAG laser lases at a fixed 1.064 micron wavelength which can be converted to either lower, visible wavelengths or higher "eyesafe" (longer than 1.4 microns) wavelengths depending on the application. Recently, research was begun at C2NVEO to efficiently generate radiation in the atmospheric transmission windows in the mid-infrared (MIR) region between 2 to 5 microns. We are investigating the use of stimulated Raman scattering (SRS), optical parametric oscillation (OPO), and combinations thereof for shifting the 1.064 micron wavelength to the MIR. SRS and OPO each have their advantages and disadvantages. SRS is a better investigated approach, can provide superb quantum conversion efficiencies, is cheap and damage resistant but suffers from a part of the energy being lost to the scattering medium. OPO has the advantage of no energy loss to the material medium but is prone to material damage and is less well developed in terms of conversion efficiencies.

This paper reports the results of SRS experiments in hydrogen gas, a first stage in a two step process. Hydrogen provides a 4155 cm^{-1} Raman shift, the largest shift of any gas. Raman scattering in hydrogen Stokes shifts the 1.064 micron Nd:YAG laser wavelength to 1.907 microns. The 1.907 micron wavelength, in turn, is a convenient staging point for further shifting to MIR wavelengths using another stage of SRS or OPO.

451

II. Experimental Setup

A. Pump Laser:

The pump laser consisted of a flashlamp pumped Q-switched Nd:YAG oscillator followed by a two stage, two pass phase conjugate amplifier. The oscillator provided 25 ns pulses at a repetition rate of 5 Hz at average powers variable from 0 to 200 mW. Two 2-mm thick etalons were
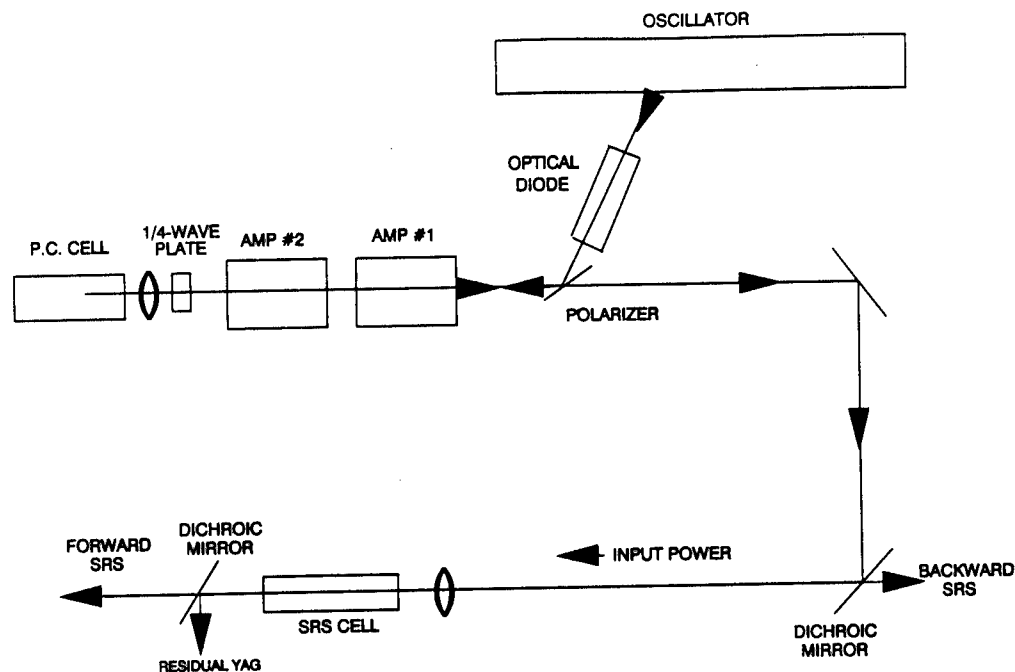

Figure 1: Experimental set up for Raman experiments.

inserted in the cavity to limit the number of longitudinal modes. Fabry-Perot analysis revealed the output contained from one to three etalon modes, depending upon the tilt of the second etalon, with approximately four longitudinal modes per etalon mode. A Faraday isolator was inserted between the oscillator and amplifier to serve as an optical diode, preventing amplified back-scattered pulsed from entering the oscillator. The amplifier rods in each amplifier were 6 mm diameter X 75 mm long. With both amplifiers fully powered, the output reached 3 watts output average power or 600 mJ per pulse. With one amplifier, the maximum output was 1.3 watts (260 mJ/pulse). Results reported below were obtained with only one amplifier turned on.

The output beam profile was measured with a CID array camera and seemed to be an Airy function at low energy extraction and a "top hat" at high power extraction. The beam waist was estimated from IR paper burns to be about 0.5 cm.

B. Raman Experiments

The Raman cell was a 40 cm long aluminum gas cell filled with hydrogen to pressures up to 1200 psig. The cell was inserted between two 45 degree Nd:YAG mirrors as shown in Figure 1. The input pump radiation at 1.064 micron was focused by an AR coated (1.06 micron) lens into the center of the Raman cell. The Raman shifted output at 1.907 micron was collinear with the pump beam and was observed as two counterpropagating beams: (1) in the direction of the pump beam propagation and (2) in the direction opposite the pump beam propagation. The former is referred to as "forward" SRS and the latter is called "backward" or "back" SRS. The input YAG pump power, transmitted YAG pump power, and SRS power were measured with two Scientech Model 365 digital power meters operating in the power averaging mode. All power readings were corrected for measured transmittance of various optics.
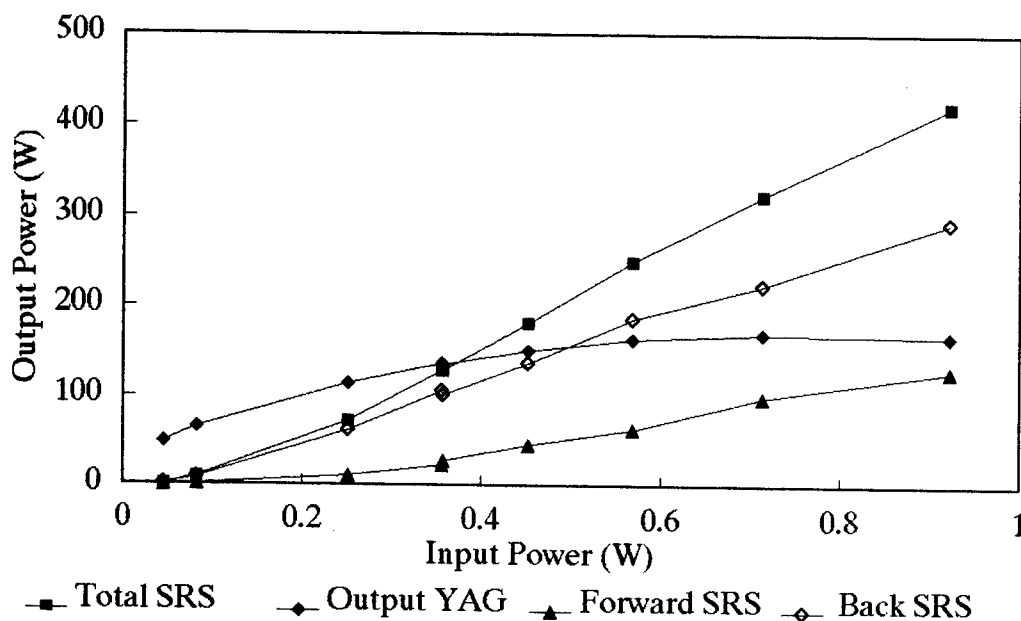

Figure 2: Output vs. input with back SRS optimized. The f-number was 48.4 and the pressure was 1080 psig.

III. Results

Figure 2 shows Raman outputs in the forward and backward directions and the residual

Nd:YAG as a function of the input power. The Raman cell was filled with 1080 psig H_2 and a 25 cm plano convex lens was used to focus the input Nd:YAG radiation within the cell. The etalons in the oscillator were adjusted for minimum forward-to-back (F/B) SRS power ratio. The back SRS was observed to have a lower threshold than forward SRS. The SRS thresholds were extrapolated from the two lowest input power data points. The thresholds determined in this manner were 39 mW for back SRS, and 45 mW for forward SRS corresponding to about 3 GW/cm^2 at the focus.
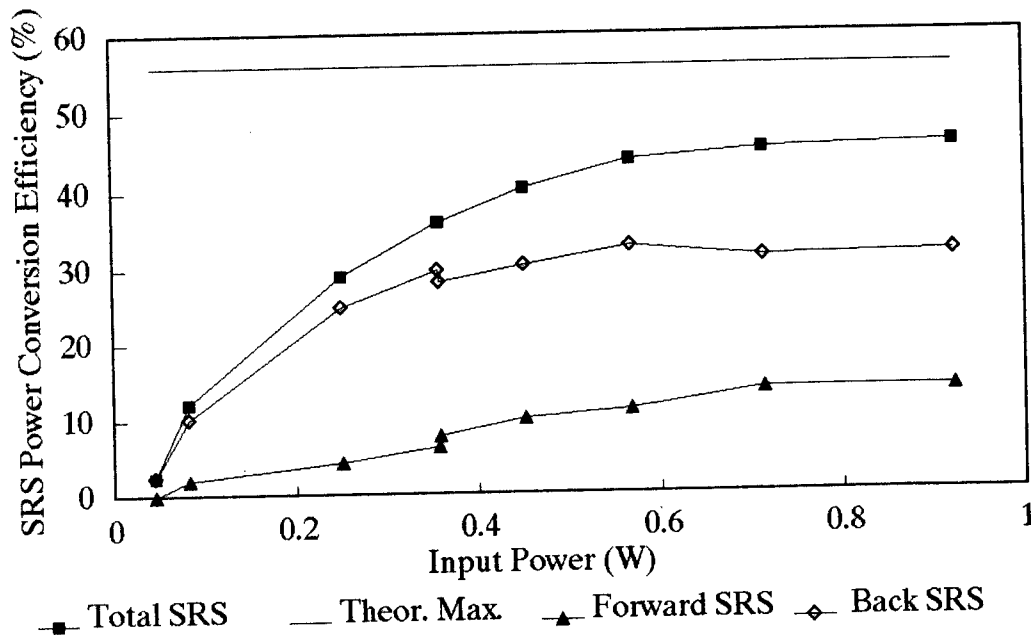

Figure 3: Raman average power conversion efficiency as a function of input power with back SRS optimized. The solid horizontal line corresponds to 100% quantum efficiency.

Figure 3 shows plots of the SRS power conversion efficiencies vs. input power. The total SRS power conversion efficiency steadily increases reaching a value of 46% at 923 mW input corresponding to a quantum efficiency of 82%. The actual conversion efficiency should be compared with the theoretical maximum energy efficiency 55.8% corresponding to 100% photon conversion of the incident pump. Figure 4 is a re-plot of Figure 3 in terms of the quantum efficiency.

The power conversion efficiency for backward SRS maximizes at 33% (59% photon conversion) at 568 mW, and decreases to 32% by 923 mW. The forward SRS power conversion efficiency increases from zero at threshold to a steady 14% (25% photon conversion) at 923 mW. The data in Figure 3 indicates that single direction efficient SRS operation may be optimized for backward SRS at lower pump powers and for forward SRS at higher pump powers.
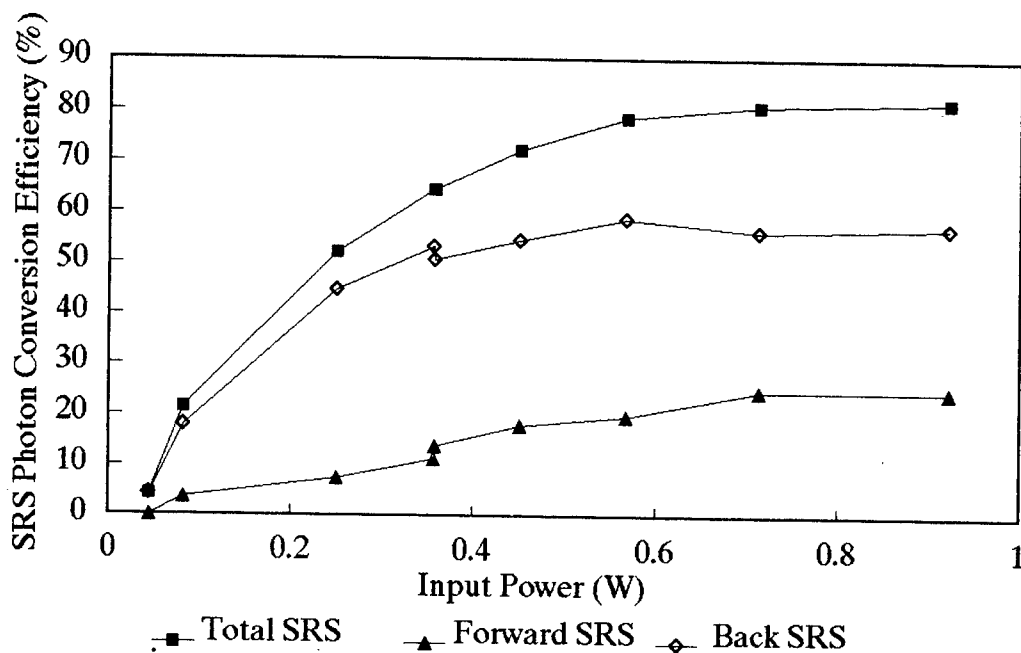
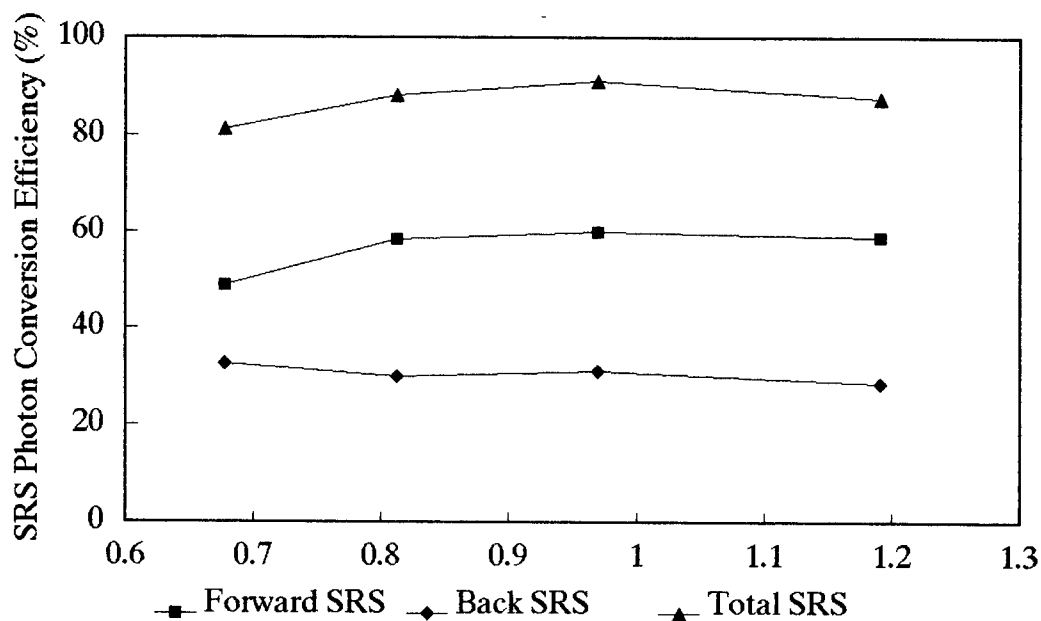Figure 4: SRS photon conversion efficiency vs. input power.

Figure 5: Photon efficiency vs. input power for forward SRS optimized. Gas pressure: 1100 psig; f-number: 58.1.

Figure 5 shows of quantum efficiencies for forward, back and total Raman conversion with forward scattering maximized. A total quantum conversion efficiency of 91% was observed at 970 mW input power.

The backward and forward SRS directions each have their advantages. The backward SRS is considered more desirable because of its phase conjugate properties. This can be especially valuable in seeding high power Raman amplifiers. For a single stage, high power conversion device, forward Stokes would be the most advantageous because F/B ratio and forward power efficiencies increase with high pump powers. One way to control the forward-to-backward (F/B) ratio is by controlling the input bandwidth.
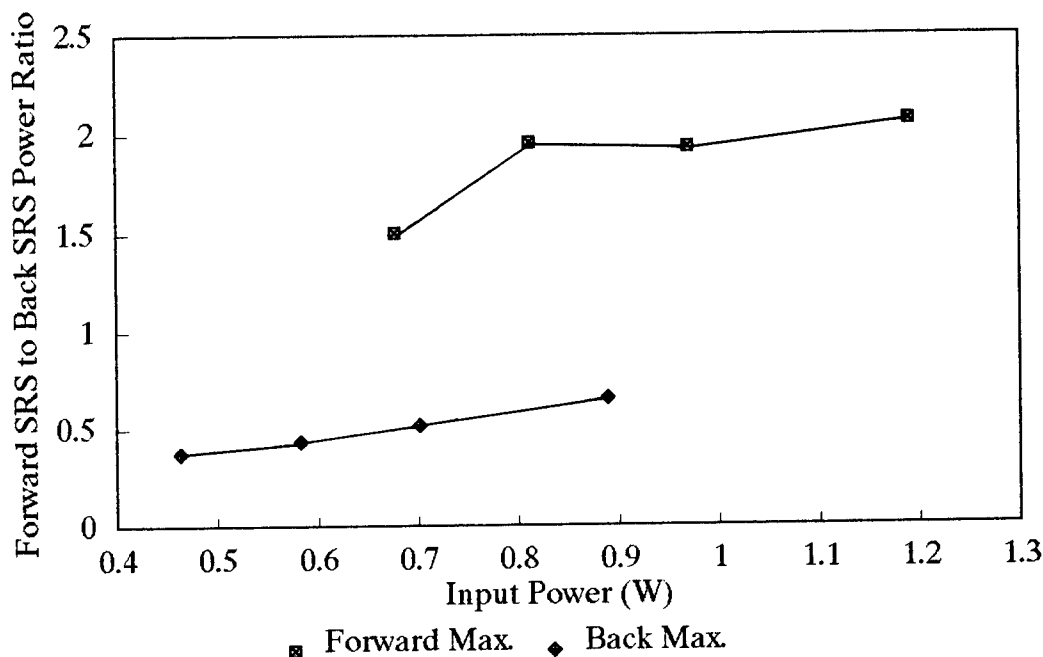

Figure 6: Mode dependent forward-to-back SRS ratio vs. input power. Gas
pressure 1100 psig, f# 58.1.

In order to determine the quantitative dependence of SRS upon the number of longitudinal modes of the pump laser, we optimized the etalons in our laser oscillator for (1) maximum F/B ratio, and (2) minimum F/B ratio. Later Fabry-Perot analysis revealed that maximum F/B corresponds to 3 etalon modes (approximately 12 longitudinal modes), and minimum F/B has 1 etalon mode (approximately 4 longitudinal modes).

Figure 6 plots the F/B ratio versus input pump power for the backward Stokes maximized (1 etalon mode) and forward Stokes maximized (3 etalon modes) cases. The F/B ratio ranges from 0.37 to 0.65 in the backward maximized case and from 1.50 to 2.07 in the forward maximized case. From Figure 6 it can be seen that backward SRS dominates at low pump powers and fewer longitudinal modes, and forward SRS dominates at high pump powers with many longitudinal

modes. The power trend of increasing F/B ratio with increasing pump powers is also apparent in both cases.

In an effort to determine effect of the lens f-number (f#) on the SRS, the f# was varied by using different focal length plano-convex lenses. The focal lengths used were 30, 25, and 20 cm. With an estimated pump beam diameter of 0.516 cm, the f#'s used were 58.1, 48.4 and 38.8 respectively. The cell pressure was 1080 psig.

Figure 7 plots the SRS output power versus f# at a constant input pump power of 1.27 W. Total and forward SRS power increased with increasing f#, and backward SRS power decreased
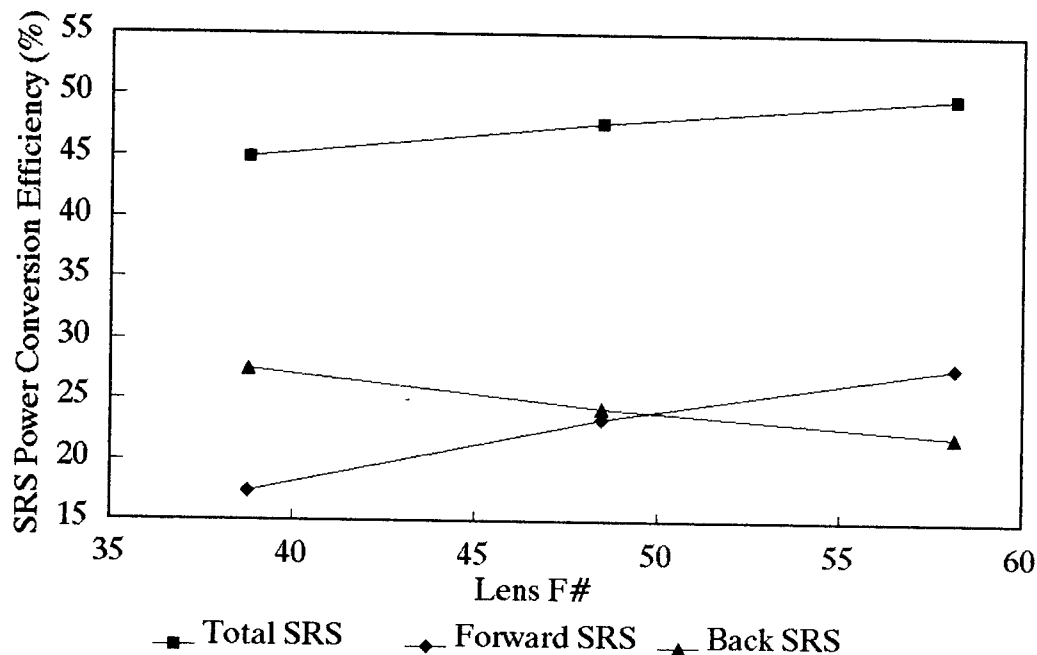

Figure 7: F-number dependent SRS power efficiency for 1.22 watt input and 1080 psig gas pressure.

with increasing f#. The same experiment was also performed at lower pump powers, and the same trends observed in Figure 7 were preserved. The total SRS power efficiency slightly increased with f#, with efficiencies (49 to 54%) approaching the theoretical maximum. Power efficiencies as high as 32% were observed for backward SRS, and 29% for forward SRS over the range of lenses used.

The dependence of SRS upon Raman cell pressure was studied with the 58.1 f# lens inserted for four pump powers, 1.22, 1.02, 0.86, and 0.72 W. Again, the trends scale with pump power, and only the data for 1.22 W is presented here. The Raman cell was filled to 1175 psig, and cell pressure was varied by bleeding off the gas.
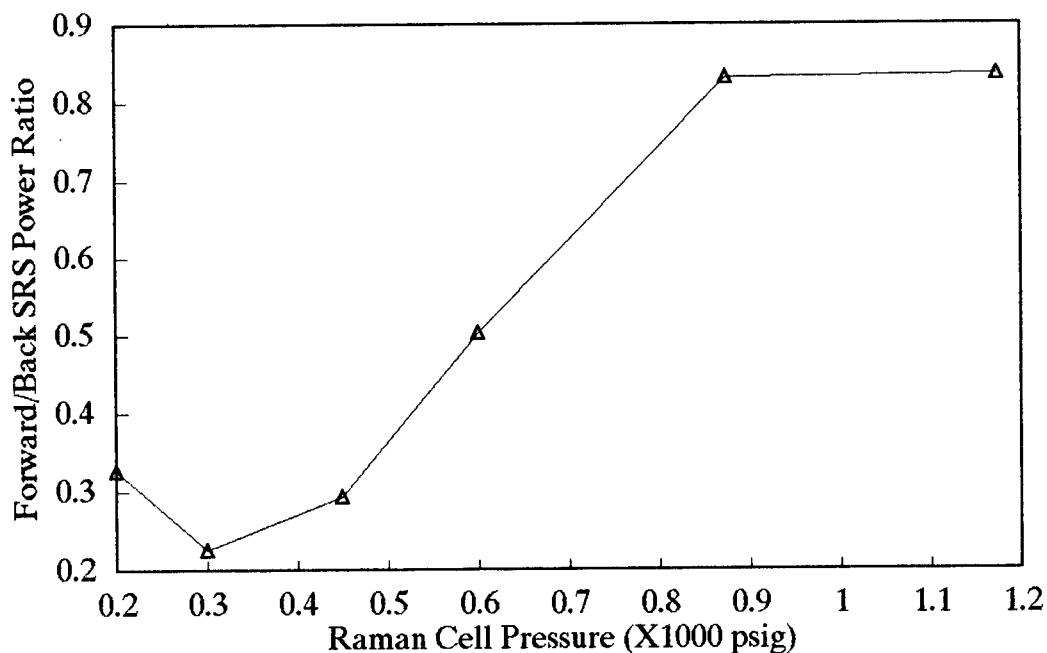
457

Figure 8: Pressure dependence of forward-to-back SRS ratio. F# 58.1, input laser power 1.18 W.
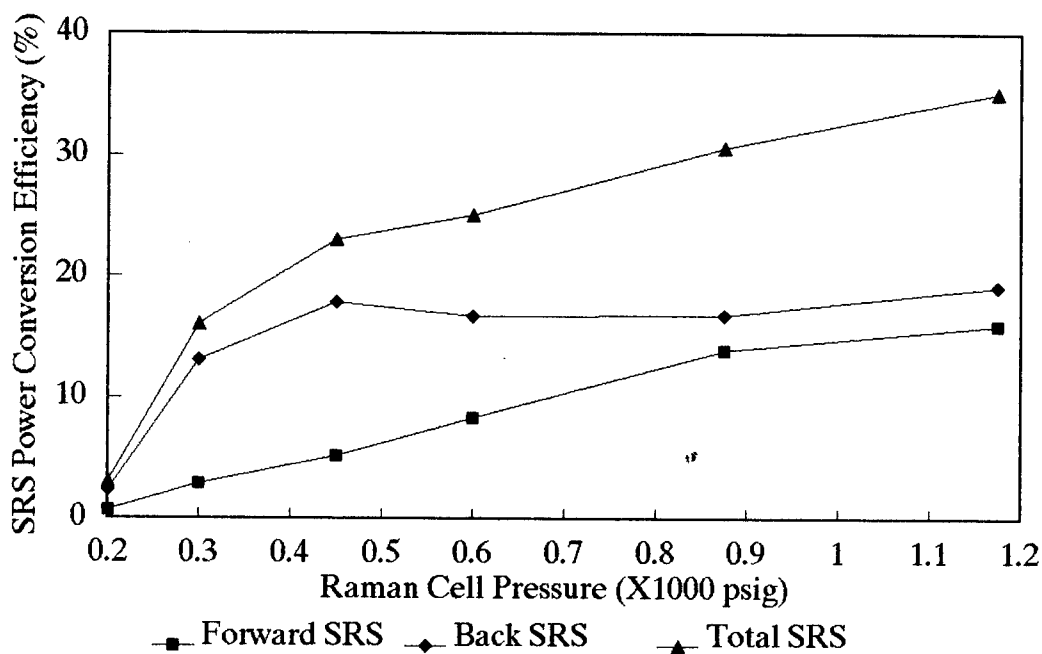

Figure 9: SRS power conversion as a function of pressure.

Figure 8 and 9 are the pressure dependences of F/B ratio and power conversion efficiencies. The F/B ratio increases with pressure, so forward SRS is optimized at high pressures while backward SRS is optimized at low pressures. Both forward SRS power and total power efficiency increase with pressure.

IV. Conclusions

Stimulated Raman scattering in hydrogen gas has been reported for quantum efficient frequency conversion of 1.064 micron to 1.907 micron. Total power efficiencies have been demonstrated in our experiments to be as high as 51%, nearing the theoretical maximum of 56%. The authors are unaware of any other experimental work in shifting 1.064 micron to 1.907 micron where such high power efficiencies were demonstrated. Total and directional SRS conversion was studied as a function of input pump power, etalon modes, f#, and cell pressure. Optimum forward SRS required high pump powers, multiple longitudinal modes, high f#, and high cell pressures. Conversely, backward SRS was optimized for low power pumping with a narrow band laser and low f#. When forward Stokes was maximized, a power conversion efficiency of 34% (60% photon conversion) into the forward Stokes was demonstrated. When backward Stokes was maximized, a power conversion efficiency of 33% (59% photon conversion) into the backward Stokes was demonstrated. As stated in the introduction, exploration of SRS in hydrogen gas was a first step in multi-step techniques to efficiently generate radiation in the 2 to 5 micron region. Further research will be done on the other steps in these nonlinear techniques. Promising future work includes degenerate and near degenerate optical parametric oscillation (OPO) of 1.907 to produce radiation around 3.8 micron. The value to the Army of generating 3.75 to 3.85 micron radiation is: (1) atmospheric transmittance is maximal in this region; (2) the atmospheric attenuation coefficient has the least worldwide seasonal variation of all IR wavelengths studied ; and (3) 3.8 micron is within the "eye-safe" spectral region. Results of this research will be published in future papers.

Hydrolysis of Toxic Organophosphorus Compounds
by Enzymes from Halophilic Bacteria

Joseph J. DeFrank, Dr.*
Tu-chen Cheng, Dr.
U.S. Army Chemical Res., Dev. & Eng. Center
Biotechnology Division, SMCCR-RSB
Aberdeen Proving Ground, MD 21010-5423

Introduction

Organophosphorus Acid (OPA) Anhydrases are enzymes that are capable of hydrolyzing a wide variety of organophosphates, including fluoride containing chemical warfare agents (soman [O-1,2,2-trimethylpropyl methylphosphonofluoridate] and sarin [O-isopropyl methylphosphonofluoridate]), acetylcholinesterase inhibitors (DFP [diisopropylfluorophosphate])and pesticides (parathion [diethyl p-nitrophenylphosphorothioate] and paraoxon [diethyl p-nitrophenylphosphate]).[1-3] In the past, these enzymes were known variously as DFPases, somanases, paraoxonases or parathion hydrolases, etc., depending upon the assay substrate used. Sources of these enzymes range from bacteria and protozoans to higher mammals including man, and the number of enzymes found has greatly increased in recent years.[3] The proliferation of both enzymes and enzyme names lead to the adoption of the name OPA Anhydrase during the First DFPase Workshop (Marine Biological Laboratory, Woods Hole, MA, June 1987) to describe these related enzymes. It was planned that this name be used until the natural substrates and functions of these enzyme are identified. Preliminary studies of OPA Anhydrases from various sources have demonstrated that these enzymes are different in terms of substrate specificity, sensitivity to inhibitors, activation by metals and molecular weight.[3] Purification and characterization of these enzymes such as the one described in this paper may assist in the determination of the true nature of their substrates, specificity and molecular structure.

The source of the enzyme to be discussed is the obligately halophilic bacterial isolate designated JD6.5. It was isolated from a warm, salt spring in 1985.[4] This isolate possesses several DFP-hydrolyzing OPA Anhydrase activities. In this paper we report on the purification and characterization of OPAA-2, the predominate enzyme from JD6.5.

Materials and Methods

Organism and cultivation: Isolate JD6.5 was obtained from Grantsville Warm Springs, which is located approximately 30 miles west of Salt Lake City, Utah, and just south of the Great Salt Lake. The primary characteristics of the springs that are of interest are a relatively constant temperature of 24° C, a pH of 6.0, and a salt content of approx. 24%.[5] Cultures were grown in a medium consisting of (grams per liter): NaCl, 50; $MgSO_2 \cdot 7H_2O$, 10; proteose peptone (Difco), 10; yeast extract, 6; and HEPES, 2.5 (pH 6.8). Inoculated flasks (4 or 6 liter) containing 1-1.5 liters of medium were incubated at 30-37° C, on a rotary shaker at 240 rpm, for 18-24 hours. Cells were then harvested by centrifugation (7,500 × g) and stored at -20° C.

461

<u>Enzyme Assays.</u> OPA Anhydrase activity was routinely assayed by monitoring fluoride release from DFP by an ion-specific electrode as has been described numerous times in the literature.[6,7] Unless otherwise stated the reaction medium contained: 500 mM NaCl, 50 mM Bis-Tris Propane, pH 7.2; 3.0 mM DFP; 1.0 mM $MnCl_2$ and 5-25 μl of enzyme sample in a total volume of 2.5 ml. Assays were routinely run at 25° C. in a thermostatically controlled vessel with stirring. The enzyme sample was preincubated in the reaction medium for 1 minute before the reaction was initiated by the addition of DFP (in isopropanol). The reaction was monitored for 4 minutes and the rate of fluoride release corrected for spontaneous DFP hydrolysis under identical conditions. One unit of OPA Anhydrase activity catalyzes the release of 1.0 μmole of F^- per minute. Specific activity is expressed as units per milligram of protein.

The hydrolysis of chromogenic substrates was conducted in an identical reaction mixture as above, with the exception of a reduction in substrate concentration to 5-100 mM. Activity was monitored by following the increase in absorbance at 405nm (for p-nitrophenol) and units expressed as 1.0 μmole of p-nitrophenol released per minute. The concentration of p-nitrophenol was determined from a standard curve with authentic material.

<u>Enzyme Purification:</u> All procedures were conducted at 4° C, and all centrifugations were at 46,000 × g for 30 minutes. Frozen or freshly harvested cells were resuspended in 10BM buffer (10 mM Bis-Tris Propane, 0.1 mM $MnCl_2$, pH 7.2) at a ratio of 3 ml of buffer for each gram of cells (wet weight). The cells were disrupted by passage through a French Pressure cell (SLM Aminco) at 16,000 psi. Cellular debris was removed by centrifugation. The crude cell supernatant , which contained the OPA Anhydrase activity, was treated with protamine sulfate to a final concentration of 2%. After centrifugation, the supernatant was fractionated with solid $(NH_4)_2SO_4$ to give the 30-65% (saturation) precipitate. The pellet was resuspended in a minimal volume of 10BM buffer and dialyzed against several changes of 20 volumes of the same buffer.

The dialyzed sample was applied to a DEAE-Sephacel (Pharmacia) column (5 × 20 cm) previously equilibrated with 10BM buffer. The column was washed to remove non-binding materials. After the washing, the elution buffer was stepped to 200 mM NaCl in 10BM. The OPAA-2 activity was eluted with a 4-liter linear gradient of 200 to 600 mM NaCl. Active fractions were pooled, concentrated by precipitation at 65% $(NH_4)_2SO_4$ and centrifugation. The pellet was dissolved in 10 mM Bis-Tris Propane, pH 7.2, supplemented with 100 mM NaCl and 5 mM KH_2PO_4. The solution was dialyzed overnight against this buffer.

The enzyme solution was loaded onto a Hydroxyapatite (HA) Ultrogel (IBF Biotechnics) column (2.6 × 14 cm) previously equilibrated with the same buffer described above. After non-binding protein was removed by washing, elution was carried out with a linear gradient of 5 to 150 mM KH_2PO_4. Enzyme fractions were again pooled and concentrated by 65% $(NH_4)_2SO_4$. After centrifugation, the pellet was redissolved in 10BM, supplemented with 10 mM NaCl, and dialyzed against the same buffer.

The enzyme solution was further purified on an HPLC-GTi system (LKB) using two GF-250 columns (0.94 × 25 cm, DuPont) in series and run with the same buffer at a flow rate of 0.5 ml/min. Fractions of 0.5 ml were collected. The pooled enzyme fractions were concentrated with a Centricon-30 concentrator (Amicon) and loaded onto a 7% polyacrylamide gel. Electrophoresis was performed according to Laemmli,[8] but without SDS and DTT. Immediately after the electrophoresis, the bands containing enzyme activity were cut out of the gel and eluted in the same electrophoresis buffer with an Extraphor Electrophoretic Concentrator (LKB).

<u>SDS-PAGE and Protein Determination:</u> Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli.[8] The protein bands were visualized with the Gelcode Silver Stain Kit (Pierce). The Coomassie Protein Assay Reagent (Pierce) was used for determination of protein concentrations with bovine serum albumin (BSA) as the standard.

<u>Preparation of polyclonal and monoclonal antibodies</u>: The polyclonal antiserum to OPAA-2 was prepared by immunizing a rat (Spraque-Dawley) with the HPLC purified enzyme fraction. Serum was obtained by tail bleeding. Spleen cells from the immunized rat were removed and fused with mouse myeloma cell line SP2/0Ag 14. The hybridoma cell clone was detected by ELISA assay using microtiter plates coated with a crude enzyme preparation. Biotinylated rabbit anti-rat IgG conjugated to horseradish peroxidase (Sigma: Rat Extravidin Staining Kit) was used for the detection of antibodies against OPAA-2. For Western blotting analysis, the same detection kit was used as in the immunobinding assay.

<u>Results</u>

<u>Organism:</u> Isolate JD6.5 is a gram-negative, aerobic, short rod, and an obligate halophile that requires at least 2% NaCl for growth (data not shown). Fatty acid analysis (Microbial ID, Newark, DE) has tentatively identified JD6.5 as a species of *Alteromonas* but not the *haloplanktis* or *putrefaciens* species that were in their database.

<u>Purification of OPA Anhydrase-2:</u> The treatment of the crude extract with 2% protamine sulfate serves to remove nucleic acids and associated proteins prior to $(NH_4)_2SO_4$ fractionation. The fractionation step (30-65%) served to concentrate the enzyme, while also giving a 3-fold purification over the supernatant with 124% recovery. This suggests that an inhibitor (or natural substrate) may have been removed during these initial steps.

When the enzyme from the $(NH_4)_2SO_4$ fractionation was applied to the DEAE-Sephacel column, it remained bound to the column while a large amount of non-binding protein passed through. As shown in Figure 1, a small peak of enzyme activity (OPAA-1) was eluted by the step to 200 mM NaCl. The bulk of the activity (>90% of the total) eluted in the linear gradient at approx. 450 mM NaCl. This chromatographic step resulted in an approx. 110-fold purification and an overall recovery of approx. 71%.

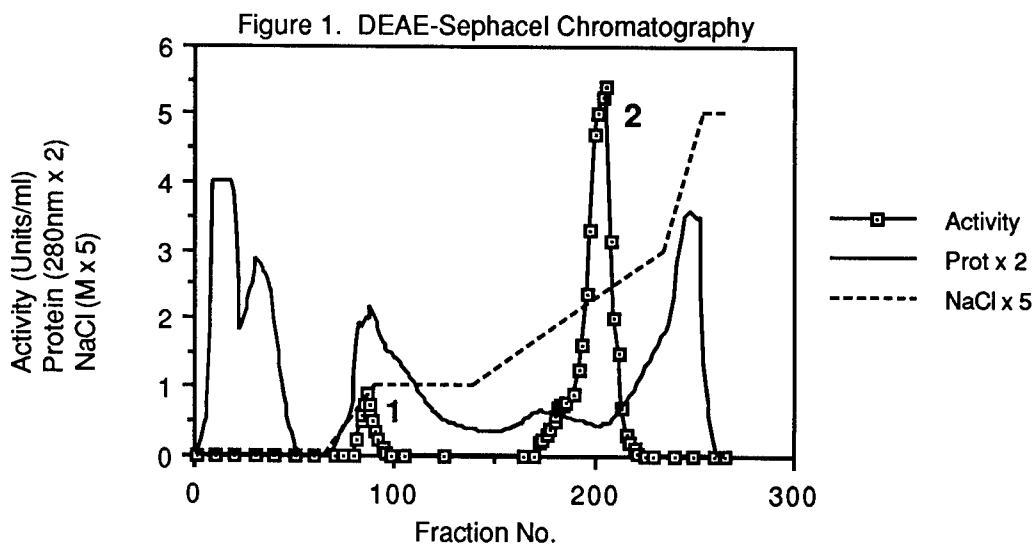Figure 1. DEAE-Sephacel Chromatography

Figure 2 shows that when the enzyme solution from the DEAE-Sephacel was applied to the Hydroxyapatite Ultrogel column, the enzyme was again retained on the column. (Note: manganese was not used during this chromatographic procedure to prevent precipitation of $MnPO_4$.) When the gradient of 5-150 mM KH_2PO_4 was passed through the column, the enzyme eluted in a single peak in an early fraction. The pooled enzyme fractions with activity over 1 unit/ml were concentrated by $(NH_4)_2SO_4$ precipitation. This column gave a 972-fold purification with an overall recovery of approx. 26%

Figure 2. HA-Ultrogel Chromatography
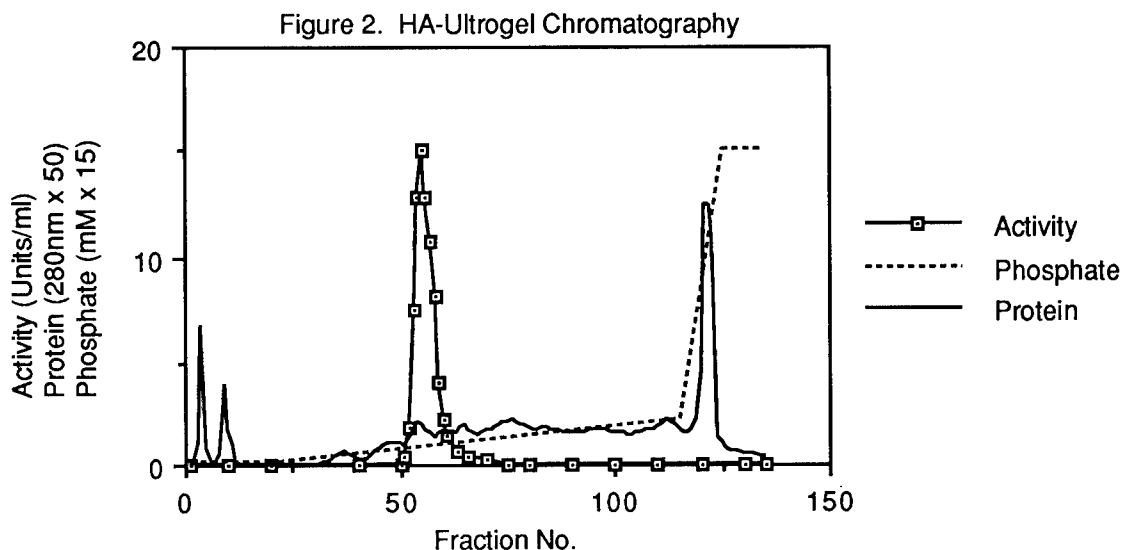
The enzyme solution was further purified through HPLC on a GF-250 (gel filtration) column in which it eluted as a sharp peak. The concentrated enzyme solution was further purified on a 7% non-denaturing polyacrylamide gel. After electrophoresis, the gel slices that contained enzyme activity were eluted by electroelution to recover the protein. The final enzyme preparation had a specific activity of approx. 300 μmoles DFP hydrolyzed/min/mg protein, representing an overall 1000-fold purification and 3% recovery. A summary of the OPAA-2 purification is shown in Table 1.

Table 1. Purification of OPAA-2 from JD6.5

Purification Step	Volume (ml)	Protein (mg)	Activity (units)	Spec. Act. (units/mg)	Purif. (fold)	Yield (%)
Crude Extract	306.0	3005.0	801.0	0.267	-	100.0
Protamine Sulfate	328.0	2938.0	853.0	0.290	1.1	106.5
Ammonium Sulfate	145.0	1167.0	992.0	0.850	3.2	123.8
DEAE-Sephacel	19.0	19.4	570.0	29.38	110.0	71.2
HA-Ultrogel	1.3	0.8	205.0	259.49	971.9	25.6
HPLC-GF 250	2.0	0.4	143.0	357.50	1337.1	17.9
PAGE	4.0	0.1	26.8	285.11	1067.8	3.3

The purified enzyme was analyzed by SDS-PAGE. A representation of the silver-stained gel is shown in Figure 3. The purification protocol described above yielded an enzyme preparation which appeared to be homogeneous as judged by a single band with a molecular weight of approx. 60,000. No difference was observed in the electrophoretical mobility of purified OPAA-2 under either reducing or non-reducing conditions. The results strongly suggest that the final purified enzyme is a single polypeptide chain.

Figure 3. SDS-PAGE Analysis of Purified OPAA-2


A - Reducing conditions

B - Non-reducing conditions
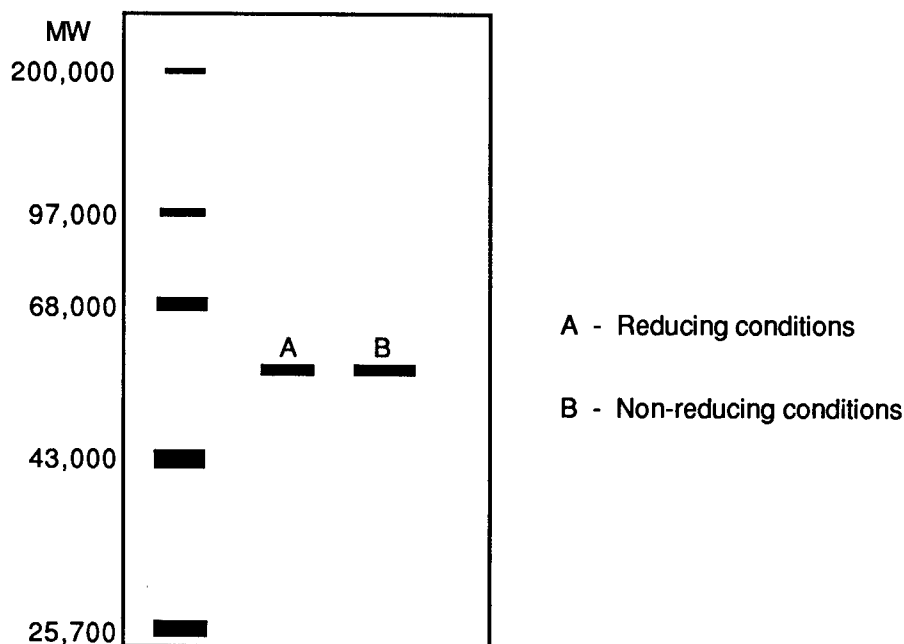
The purified OPAA-2, in addition to the pooled fractions 1 and 2 from the DEAE-Sephacel step and crude extract, were also analyzed by Western blotting after SDS-PAGE. The blots were analyzed with either monoclonal antibody, MAb#6 (Fig. 4A) or polyclonal anti-serum (Fig. 4B). A single band of purified OPAA-2 was detected on both blots (Lane 4). The monoclonal antibody was shown to react with two protein bands (MW 78,000 and 74,000) in both the crude extract and pooled DEAE fraction 1 (A-1 and A-2). In the blot reacted with anti-serum, these two bands and a third band with an estimated molecular weight of 53,000 were also detected (B-1 and B-2). The results suggest that these three protein bands are separated from OPAA-2 during the DEAE-Sephacel chromatography.To characterize the anti-serum in terms of biological activity, the enzyme activity of purified OPAA-2 after immunoprecipitation was determined. At a 250-fold dilution of anti-serum, a 50% inhibition of enzyme activity was observed (data not shown).

Substrate Specificity: The specific activity of OPAA-2 against a variety of substrates is summarized in Table 2. Of all the substrates tested, the highest activity was with soman and DFP. The enzyme also exhibited activity against sarin and two chromogenic compounds, p-nitrophenyl-ethyl(phenyl)-phosphinate (NPEPP) and p-nitrophenyl-methyl(phenyl)phosphinate (NPMPP)(data not shown). Preliminary NMR analysis has shown that all isomers of soman and sarin are hydrolyzed at equivalent

rates. Paraoxon was hydrolyzed by some preparations of the purified enzyme at about 10% the rate of DFP.

Figure 4. Western Analysis of JD6.5 Preparations


A - Monoclonal

B - Polyclonal
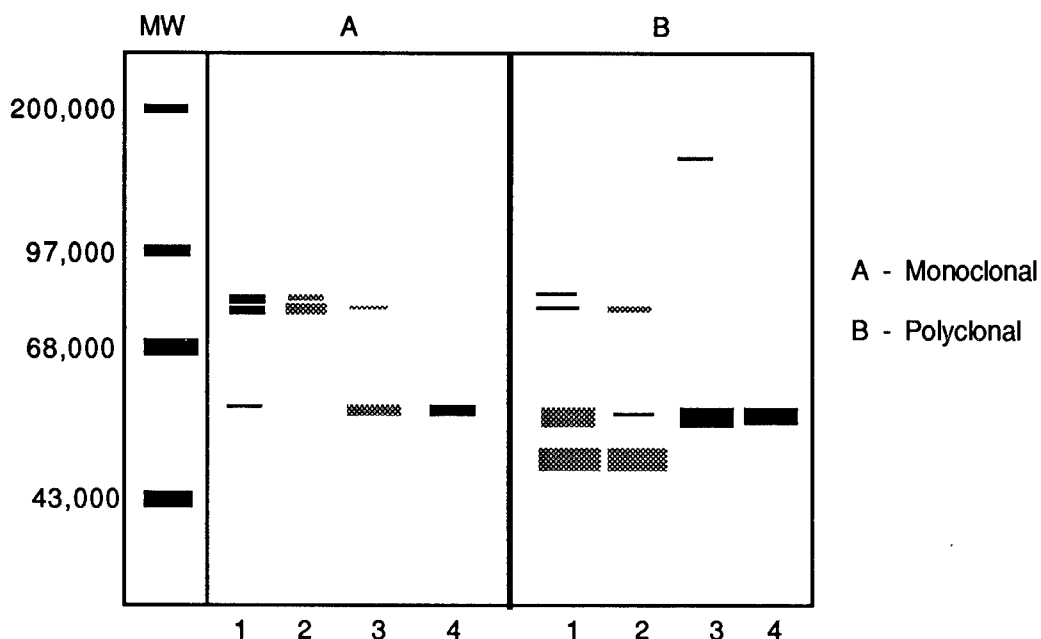
In addition to these compounds, a variety of potential substrates for esterases, phosphatases, phosphodiesterases, phosphotriesterases and phospholipases were examined as potential substrates. These compounds showed little or no activity with OPAA-2. Only Mipafox (N,N'-diisopropyl phosphorodiamidofluoridate), which is only slightly hydrolyzed by OPAA-2, exhibited a high level of inhibition of the enzyme for DFP hydrolysis. Figure 5 shows the effect of incubation of OPAA-2 with Mipafox. In this assay, the Mipafox was added to the reaction mixture during the preincubation period before the addition of DFP. At 3.0 μM Mipafox, the DFP hydrolysis was inhibited to greater than 90%. This inhibition is completely reversible. Dialysis of the inhibited enzyme overnight against 10BM buffer containing 100 mM NaCl resulted in complete restoration of activity against DFP (data not shown).

Table 2. Activity of OPAA-2 (μmoles/min/mg)

| Substrate	[mM]	pH 7.2		pH 8.5	
		(-Mn)	(+Mn)	(-Mn)	(+Mn)
Soman (GD)*	10.0	173.20		181.10	
DFP	3.0	72.96	113.74	87.19	143.60
Sarin (GB)*	10.0	15.40		50.10	
NPEPP	0.1	14.14	26.49	13.23	20.74

* Activity determined by NMR analysis

466

Figure 5. Effect of Mipafox on OPAA-2
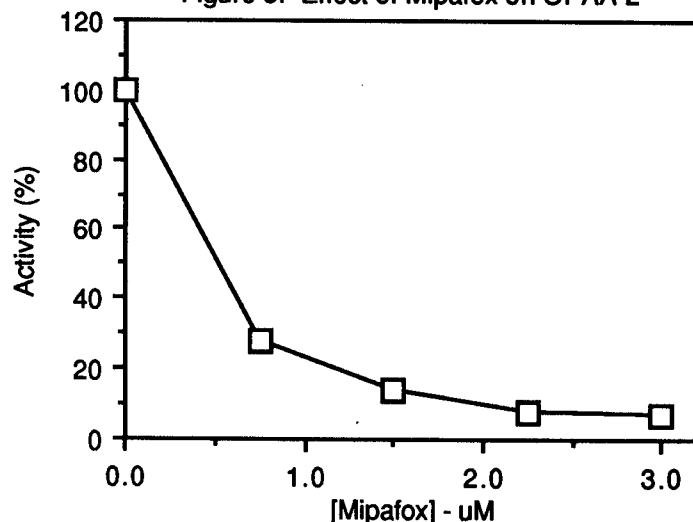
Effect of pH and Temperature: With DFP as substrate, the effect of pH on the activity of OPAA-2 was examined. In addition, the kinetic parameters were determined at each pH value. The results are shown in Figure 6. The K_{cat} is the catalytic constant or turnover number for the enzyme. It represents the number of substrate molecules converted to products per active site per second and was calculated by multiplying the specific activity of the enzyme by its molecular weight in thousands and dividing the result by 60. Since the molecular weight of OPAA-2 is 60,000, the numerical value for K_{cat} (s^{-1}) is the same as its specific activity.

Figure 6. pH Profile of OPAA-2
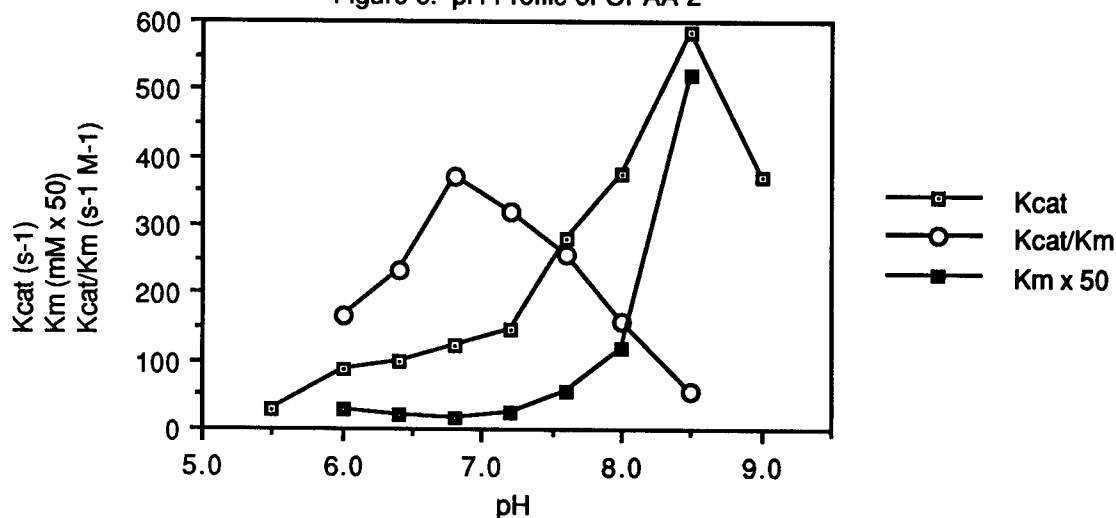
The K_M is the dissociation constant of the enzyme/substrate complex and is plotted as [mM]. The parameter K_{cat}/K_M is sometimes referred to as the "specificity constant" or "catalytic efficiency". The units for K_{cat}/K_M are $(s^{-1}M^{-1})$. Since the diffusion-controlled encounter frequency of an enzyme and a substrate should be about $10^9 \ s^{-1}M^{-1}$, this represents the upper limit for K_{cat}/K_M. A "perfect" enzyme should have a K_{cat}/K_M of around 10^8 to 10^9 and a K_M significantly greater than the concentration of substrate normally encountered. As shown in Figure 6, the observed pH optimum for OPAA-2 was found to be 8.5. The highest level of catalytic efficiency (K_{cat}/K_M) was seen at pH 6.8. This value reflects the pK_a of the enzyme and the identity of ionizing groups in the active site or on the surface. From these observations, the presence of a histidine or a cysteine is suggested.

The effect of temperature on the reaction rate of the hydrolysis of DFP by OPAA-2 is shown in Figure 7. The initial reaction rate of the enzyme with 3.0 mM DFP reached its maximum at 50° C at pH 7.2 in the presence of 1.0 mM $MnCl_2$. The level of activity decreased at 55° C which was the highest temperature tested in order to protect the fluoride electrode. The purified enzyme could be stored for months at -70° C in the presence of dithiothreitol (DTT). The addition of DTT not only improved stability but appeared to stimulate activity over the course of six months (data not shown). To confirm the stimulation of OPAA-2 activity by reducing agents, enzyme samples were incubated at 4° C in the presence of 0.1 or 1.0 mM DTT or ß-mercaptoethanol (BME). Both reducing agents showed a stimulatory effect, with DTT causing the greatest increase (Figure 8).

Figure 7. Temperature Profile of OPAA-2
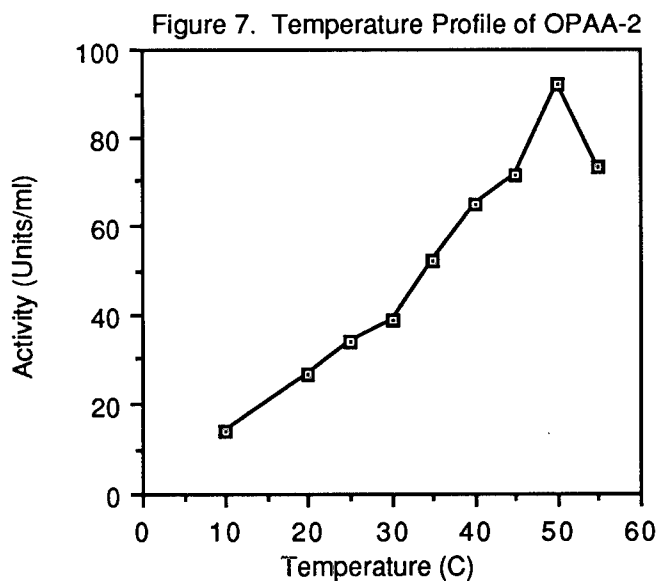
Effect of Sulfhydryl Inhibitors: The effect of several sulfhydryl-specific inhibitors was examined. The inhibitors, p-chloromercuribenzoate (PCMB), iodoacetic acid (IAA) and N-ethylmaleimide (NEM) were added to the enzyme at 0.1, 1.0 and 5.0 mM and preincubated for 60 minutes at 4° C. As illustrated in Figure 9, all these reagents caused an inhibition of OPAA-2 activity against DFP. The results suggest that the enzyme requires a functional sulfhydryl group for its activity.

<u>Effect of Metals</u>: The effects of various metal ions on OPAA-2 activity against DFP were conducted in the following manner. The enzyme sample was initially tested, as is, with the various metals shown. These results are given by the first bar in the clusters in Figures 10(A,B,C). An identical sample of enzyme was incubated with 0.1 mM EGTA (ethylenebis(oxyethylenenitrilo)tetraacetic acid) for 60 minutes at room temperature, then assayed for activity with DFP.

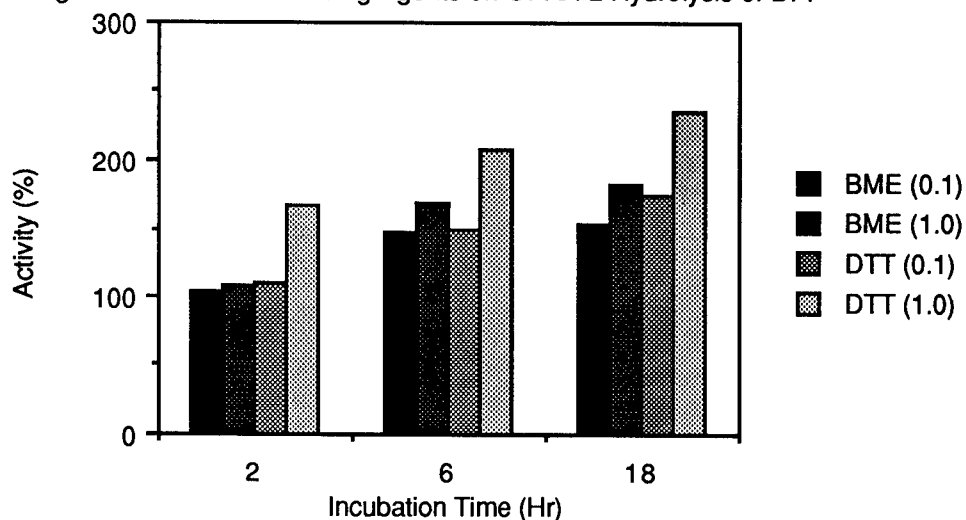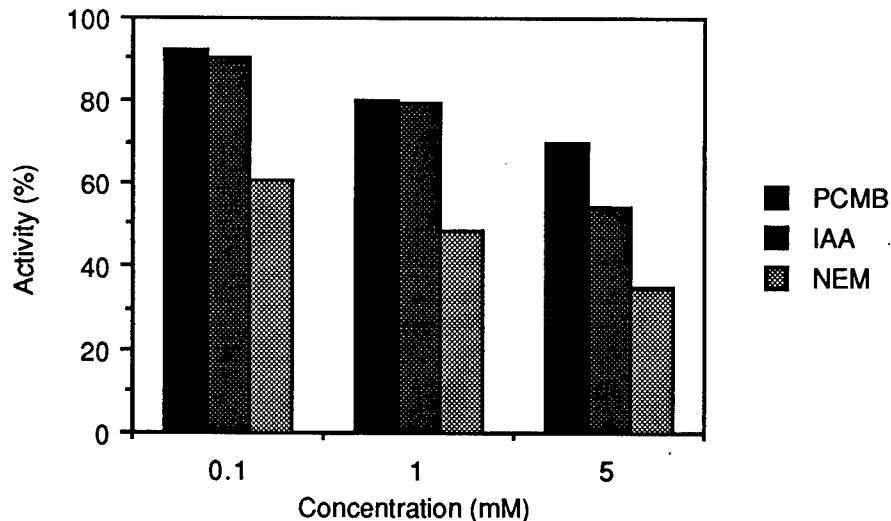Figure 8. Effect of Reducing Agents on OPAA-2 Hydrolysis of DFP

Figure 9. Inhibition of OPAA-2 by Sulfhydryl Reagents

The EGTA-inhibited activity is represented by the second bar in each cluster. The third bar shows the activity upon addition of 0.4 mM of the metal being tested. A preincubation of 2 minutes was

conducted before the activity was measured. The fourth bar of the cluster shows the effect of the addition of 0.4 mM MnCl$_2$ to the previous reaction mixture to determine if any addition stimulation could be obtained.

Figure 10A. Metals/EGTA Treatment - Group 1
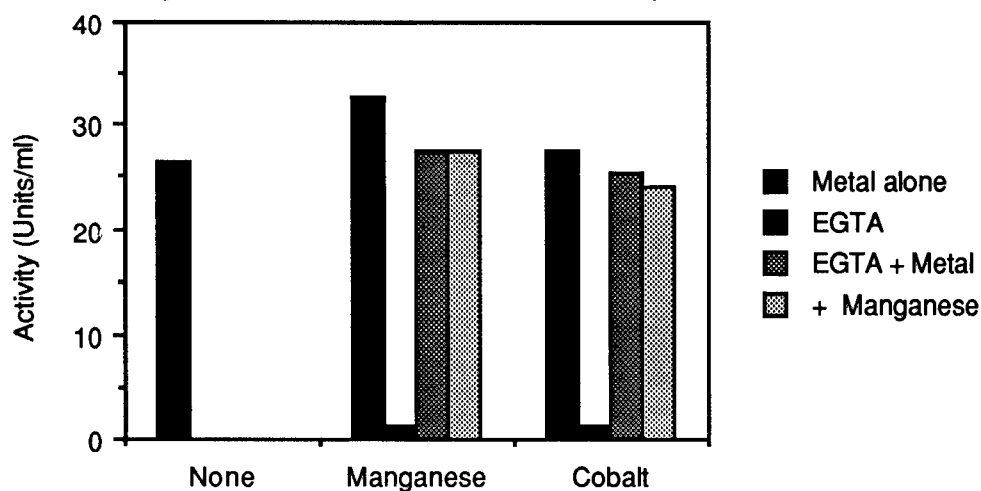
Figure 10B. Metals/EGTA Treatment - Group 2
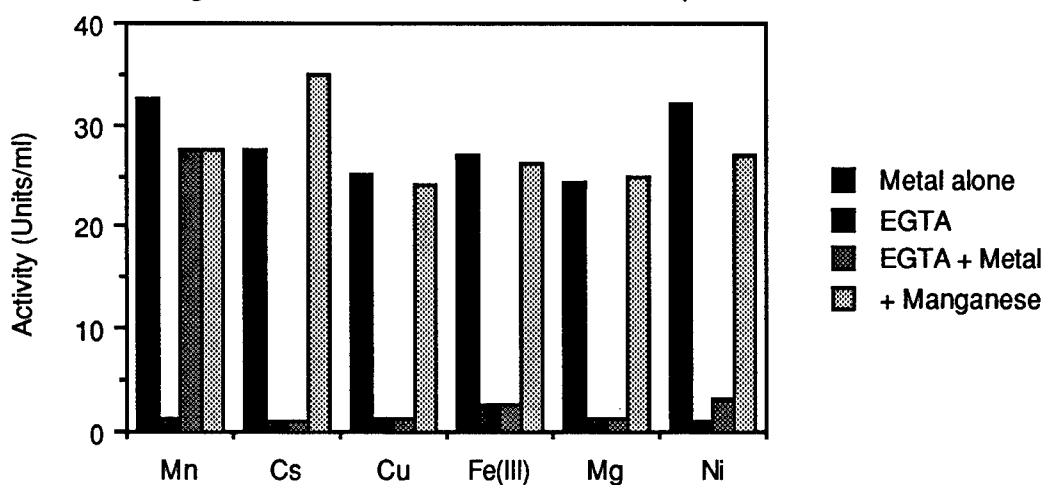
Three types of reactions were observed. In the first type, both manganese and cobalt showed nearly complete reactivation of the enzyme activity with no additional stimulation by further manganese addition (Fig. 10A). The second group of metals (cesium, copper, iron, magnesium and nickel) showed little or no reactivation of the enzyme, but also did not block reactivation following manganese addition (Fig. 10B). The third group of metals (calcium and zinc) not only inhibited the enzyme to varying degrees, but also prevented reactivation by manganese (Fig. 10C). Zinc, in

particular, showed significant inhibitory effect against OPAA-2, both as an exogenous inhibitor as well as blocking its reactivation. In a number of instances, the reactivation of the enzyme by manganese following EGTA treatment resulted in much higher activity than that observed in the original sample. The above results would suggest that the native enzyme may have contained significant levels of calcium or zinc that rendered it inactive. The possible effect of the addition of manganese or other metals to the growth media has yet to be investigated.

Figure 10C. Metals/EGTA Treatment - Group 3
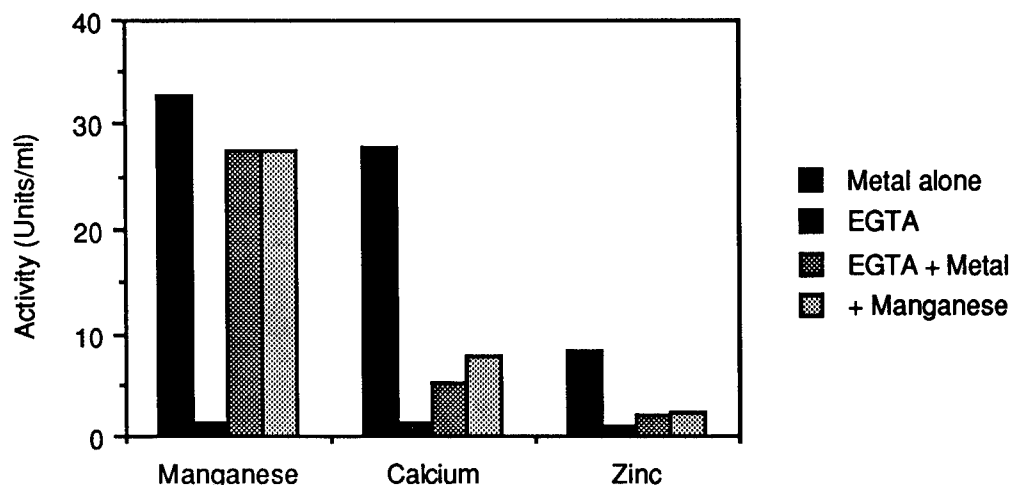
Discussion

Halophilic isolate JD6.5 possesses several OPA Anhydrase activities. Among all the OPA Anhydrases described from various sources, the enzyme from JD6.5 has been shown to have the highest activity. The predominant enzyme, OPAA-2, can be separated from other OPA Anhydrases by DEAE-Sephacel chromatography. Using DEAE-Sephacel at pH 5.0, OPAA-1 has been found to contain at least two DFP hydrolyzing activities (data not shown).

By use of the procedures described in this paper, OPAA-2 has been purified to homogeneity with a purification factor of greater than 1000-fold. SDS-PAGE indicates that the enzyme is composed of a single polypeptide with a molecular weight of 60,000. This value corresponds well with the apparent molecular weight obtained in preliminary studies using Sephacryl S-200, gel permeation chromatography (data not shown). Since several attempts to purify OPAA-2 by hydrophobic interaction chromatography on Phenyl Sepharose CL-4B were unsuccessful due to the tight binding of the enzyme to the column, it can be inferred that this is a very hydrophobic protein

The purified enzyme has been shown to possess high activity against several toxic organophosphorus compounds including DFP, paraoxon, soman, and sarin. Studies with the purified enzyme demonstrated that OPAA-2 has a pH optimum of approx. 8.5 and a temperature optimum of 50° C. Under these conditions the estimated turnover number for the enzyme, with DFP as substrate, would be approx. 100,000 molecules/min/enzyme molecule. At room temperature, the turnover number is still a relatively high 35,000 min^{-1}.

The enzyme is strongly, but reversibly, inhibited by both EGTA and Mipafox. Stimulation and stabilization of the enzyme is observed with reducing agents such as DTT or BME, and by manganese or

cobalt. The requirement for a sulfhydryl group was demonstrated by the inhibition of enzyme activity by PCMB, IAA and NEM. This is also in line with the possible presence of a active sulfhydryl group (cysteine) suggested by the pH studies. However, these results do not eliminate the possibility of the involvement of a sulfhydryl group at a location distant from the active site, but important for the three dimensional configuration of the protein or its stability.

The inhibitory effect of EGTA was examined owing to its stronger binding of divalent cations than EDTA. At 0.1 mM EGTA a nearly complete inhibition of OPAA-2 activity against DFP was observed. Similar attempts to inhibit the enzyme with EDTA showed only a slight effect (data not shown). Surprisingly, upon the addition of manganese to the EGTA treated enzyme, the activity occasionally rose to levels considerably higher than with the original enzyme sample. This effect varied from one preparation of enzyme to another. A possible explanation of this phenomenon is that the native enzyme, when synthesized, becomes partially loaded with metal ions such as calcium or zinc. which are not optimal for activity. The removal of these metals by the EGTA would allow complete loading with manganese and maximum enzyme activity.

Polyclonal anti-serum and monoclonal antibodies have been prepared against the purified OPAA-2. The characterization of OPAA-2 by Western blotting analysis has been carried out with these antibodies. Additional preliminary results have shown that these antibodies react with proteins in other halophilic and thermophilic bacteria that have been previously shown to have DFP hydrolyzing OPA Anhydrases. The results suggest that these functionally related enzymes may share common sequences, but does not eliminate the possibility of reactivity with completely unrelated proteins.

Acknowledgements

References

1. Mazur, A. 1946. An Enzyme in Animal Tissue Capable of Hydrolyzing the Phosphorus-Fluorine Bond of Alkyl Fluorophosphates. *J. Biol. Chem.*, **164**: 271-289.

2. Mounter, L.A. 1963. Metabolism of Organophosphorus Anticholinesterase Agents. In *Hanbuc de Experimentellen Pharmakologie: Cholinesterases and Anticholinesterase Agents*, Kolle, G.B., ed., Springer-Verlag, Berlin, 486-504.

3. Landis, W.G. and DeFrank, J.J. 1990. Enzymatic Hydrolysis of Toxic Organofluorophosphate Compounds. In *Biotechnology and Biodegradation, Proceedings of the International Workshop on Biotechnology and Biodegradation (1989), Lisbon, Portugal, Advances in Applied Biotechnology Series, Volume 4.* Kamely, D., Chakrabarty, A. and Omenn, G.S., eds., Gulf Publishing Company, Houston, 183-201.

4. DeFrank, J.J. and Cheng, T.-c. 1989. Characterization of a Diisopropylfluorophosphate-Hydrolyzing Enzyme from a Halophilic Bacterium. *Abst. Annu. Meet. Am. Soc. Microbiol.*, **K186**:276.

5. Mundorff, J.C. 1970. Major Thermal Springs of Utah. *Water-Resources Bulletin 13*, Utah Geological and Mineralogical Survey and U.S. Geological Survey.

6. Hoskin, F.C.G. and Roush, A.H. 1982. Hydrolysis of Nerve Gas by Squid Type Diisopropylphosphorofluoridate Hydrolyzing Enzyme on Agarose Resin. *Science*, **215**:1255-1257.

7. Landis, W.G., Haley, M.V. and Johnson, D.W. 1986. Kinetics of the DFPase Activity in *Tetrahymena thermophila. J. Protozool.*, **33**:216-218.

8. Laemmli, U. 1970. Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4. *Nature* (*London*), **227**:680-685.

SHALLOW, LOW TEMPERATURE Au-Ge OHMIC CONTACTS TO GaAs (U)

*Michelle A. Dornath-Mohr, Ms
Melanie W. Cole, Ms
Donald W. Eckart, Mr.
Richard T. Lareau, Dr.
Wayne H. Chang, Mr.
Kenneth A. Jones, Dr.
U.S.Army ETDL, Fort Monmouth, NJ 07703

INTRODUCTION

As semiconductor devices approach submicron dimensions, ohmic contacts take on increased importance as the percent of the parasitic resistance attributeable to them increases. In addition, component isolation of discrete devices becomes more difficult and retaining superlattice integrity in heterostructure devices is more challenging.

The source-drain resistance in field effect transistors (FET) is one of the most important parameters in determining ultimate device speed. The source and drain contact resistance contributes to this resistance, with their percentage contribution increasing as the channel length is shortened. Therefore, reduction of the contact resistance will increase the ultimate speed of the device. In a dual channel high electron mobility transistor (DC-HEMT) electrical contact must be made with one channel without simultaneously contacting the second channel which is about 50 nm away. Therefore, a shallow contact with a smooth interface and a low contact resistance is required for this device to operate properly.

Non-alloyed contacts have the potential to be shallow with smooth interfaces because they are formed by a diffusion process as opposed to the melting process of alloyed contacts. Diffusion is a slower process and therefore more easily controlled. This allows the process to be studied in more detail and therefore a better understanding of the fundamental mechanisms can be obtained. This could lead to the design of better contacts.

Although numerous studies of Au-Ge contacts have been made [1-13], most of them have focused on the alloyed contact [1-10] which requires annealing the samples above the 356°C Au-Ge eutectic temperature. Studies that have dealt with non-alloyed contacts have investigated the electrical properties [11-13], but little work has been done to correlate the electrical measurements with metallurgical studies.

In addition to providing shallow, smooth ohmic contacts, low temperature processing allows the fundamental mechanisms of ohmic contact formation to be studied. In non-alloyed contacts the formation steps occur sequentially over an extended period of time as opposed to the almost instantaneous reactions during the formation of a typical alloyed contact.

In this paper we examine two proposed mechanisms for ohmic contact formation. One hypothesis is that an ohmic contact is formed by heavy As doping of the Ge in intimate contact with the GaAs[14]. The other hypothesis is that heavy Ge doping of the GaAs[4,15] allows considerable tunneling at the metal-semiconductor interface. Barrier height reduction [16,17], segregation of the Au and Ge[18], evaporation of the As[19], formation of Au-Ga phases[17,20,21], and anomalous diffusion at the interface[22] must all be considered.

This study correlates transmission line model (TLM) and Schottky barrier height measurements with scanning electron microscopy/energy dispersive spectroscopy (SEM/EDS), transmission electron microscopy (TEM), Rutherford backscattering spectroscopy (RBS), Auger electron spectroscopy (AES), and secondary ion mass spectroscopy (SIMS).

EXPERIMENTAL

The substrates used for TLM and Schottky barrier height measurements were (100) grown GaAs wafers doped with Si to 10^{17} at/cm^3. The substrates used for the material characterization techniques were (100) semi-insulating GaAs wafers. Following an organic degrease step, the wafers were patterned using standard lithographic techniques. Mesas were defined by an etch of H_2SO_4:H_2O_2:DI (1:8:100). Prior to metallization, the wafers were etched in a solution of NH_4OH:DI (1:100) to remove the native oxide layer. Au-Ge films with compositions of 10, 27 (eutectic composition), 50, and 75 atomic percent Ge were formed by sequential electron beam evaporation of the specified amount of Au and Ge (two-layer structure). Deposition rates were approximately 3 Å/sec. The total thickness of the metallization was held constant at 75 nm. Films on a second set of samples were formed by depositing 4 sequences of 15 nm Au followed by 7.5 nm Ge (multi-layer structure). To study the effect of an

encapsulation layer, a 100 nm Si_3N_4 layer was deposited by PECVD at 300°C for 12 min. on some of the metal films before annealing.

Short anneals ($\leq$ 10 min.) were carried out in a nitrogen ambient using a commercial RTA system (HEATPULSE - 410) and longer anneals were performed in an argon ambient using a conventional furnace. The temperature was held constant at 320°C while the time was varied from 5 sec to 33 hr. The 320°C temperature was chosen to insure the annealing temperature stayed well below the Au-Ge (356°C) and Au-Ga (341°C) eutectics, thus forming no liquid. After annealing the Schottky barrier height samples, a mesa etch was performed to eliminate the leakage current around the periphery of the diodes. The Si_3N_4 was removed from the capped samples after annealing by a quick dip in buffered HF.

An AMRAY 1610 SEM with EDAX PV 9100 EDS system was used for surface analysis. A PHI 660 scanning Auger microprobe/scanning electron microscope equipped with Ar^+ ion sputtering was utilized to examine several contact structures. SIMS analyses were obtained using a Cameca IMS-3f magnetic-sector and a PHI 6300 quadrupole based ion microanalyzers with primary ion beams of oxygen, at an impact energy of 8 KeV over approximately 250 um x 250 um rastered area, and with cesium, at an impact energy of 5.5 KeV over approximately 50 um x 50 um rastered area. An analyzed area of approximately 10% of the total sputtered region was accepted for secondary ion monitoring. RBS measurements were performed using a General Ionex Tandetron 4117A High Voltage Ion Accelerator with a 3.0 MeV He^{++} ion beam. The (110) cross-sectional TEM samples were prepared using standard sample preparation techniques. The TEM study was performed on a Philips 420T STEM operated at 120 KV. The high resolution electron microscopy was performed on an ISI 002B TEM operated at 200 KV. The microprobe with a beam size of 5 nm was used for EDS microanalysis on a PGT EDS system. An EDAX PV 9100 system was also used with a probe size of 30 nm. Phase identification was achieved via selected area diffraction (SAD) and microdiffraction methods.

RESULTS

The mixing of the Au and Ge occurred more rapidly in the multi-layer structures than in the two-layer structures. The surface morphology varied slightly between the two structures although no quantitative differences could be found. Therefore, we will not distinguish between these two types of structures in the following discussions.

TLM measurements were used for ohmic contact evaluation. The I-V curves for the standard ohmic contact samples (uncapped, 27 at % Ge) in Figure 1 show that electrical changes occur after short anneals. They

stabilize after approximately 5 min. of annealing and show ohmic behavior after a 3 hr. anneal. The samples annealed for 3 hr. have a specific contact resistance, r_c, of approximately 5×10^{-6} ohm-cm^2. Further annealing up to 9 hr. has little effect on r_c. The capped sample is also ohmic after a 3 hr. anneal, but its contact resistance is much higher at 4×10^{-4} ohm-cm^2. Specific contact resistance also varies dramatically with changes in the Ge concentration. Samples with 10 at. % Ge require a 6 hr. anneal to show ohmic behavior with $r_c = 8 \times 10^{-6}$ ohm-cm^2. The 50 at. % Ge samples require a 9 hr. anneal before exibiting ohmic behavior with $r_c = 1 \times 10^{-2}$ ohm-cm^2. The 75 at. % Ge samples are not ohmic even after a 33 hr. anneal.
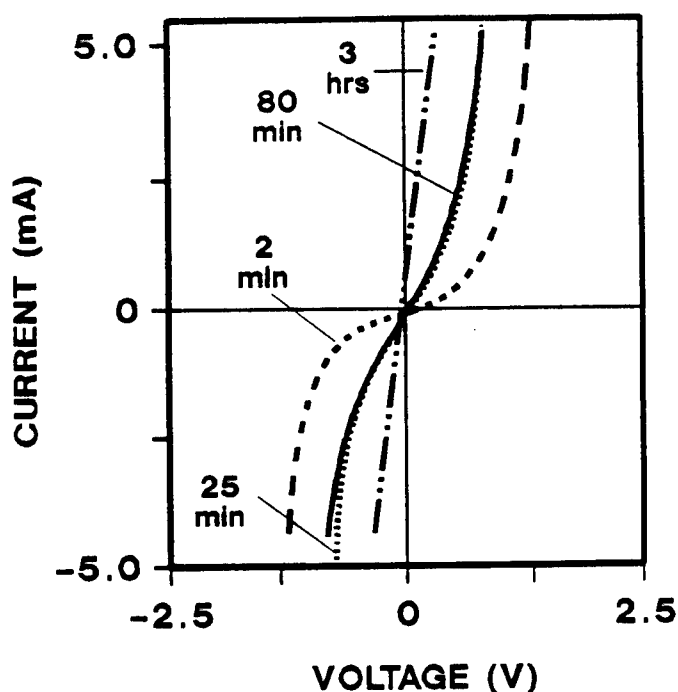
Figure 1. IV curves as determined by TLM measurements for samples annealed for 2, 25, 80, and 180 min.

Schottky barrier heights are plotted as a function of annealing time in Figure 2. Most of the decrease in the barrier height, from 0.75 to 0.40 eV, occurs during the first 5 min. of annealing. This corresponds to the interval of time over which the initial change in the TLM measurements occured. The.capped samples started with reduced barrier heights as compared with the uncapped samples. This can be attributed to the heating of the substrate to 300°C during the nitriding process, implying that some interdiffusion of the metals occurred. The fact that the barrier heights for the capped and uncapped samples converge after 180 sec. suggest that the barrier height reduction is not appreciably affected by the cap.
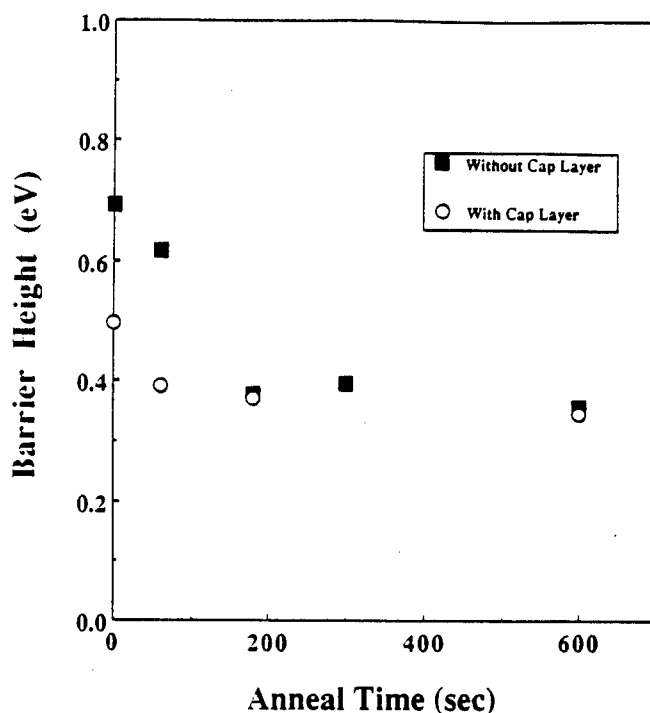
Figure 2. Barrier heights determined electrically, plotted as a function of annealing time for capped and uncapped samples.

The SEM micrograph in Figure 3a shows the segregation of the metals into Au and Ge rich regions that occurs within only 5 sec of annealing. Au-Ge segregation seems to be almost complete following a 240 sec. anneal (Figure 3b) since no significant changes occur in the surface morphology between the 240 sec. and 3 hr. anneals (Figure 3c). With a Si_3N_4 cap (Figure 3d), the Au rich and Ge rich regions are much smaller after a 3 hr. anneal than in the uncapped sample.

Figure 3. SEM micrographs of a.) a sample annealed for 5 sec., b.) a sample annealed for 240 sec., c.) a sample annealed for 3 hr., and d.) Si_3N_4 capped sample annealed for 3 hr.

AES maps confirm the presence of micron size Au and Ge rich regions. AES sputter depth profiles for Au and Ge rich regions do not reveal Ga or As presence near the surface (signals below Limit-of-Detection (LOD) 0.1%) for either the Au rich or the Ge rich regions. SIMS ion images in Figure 4 show that the Au and Ge tend to segregat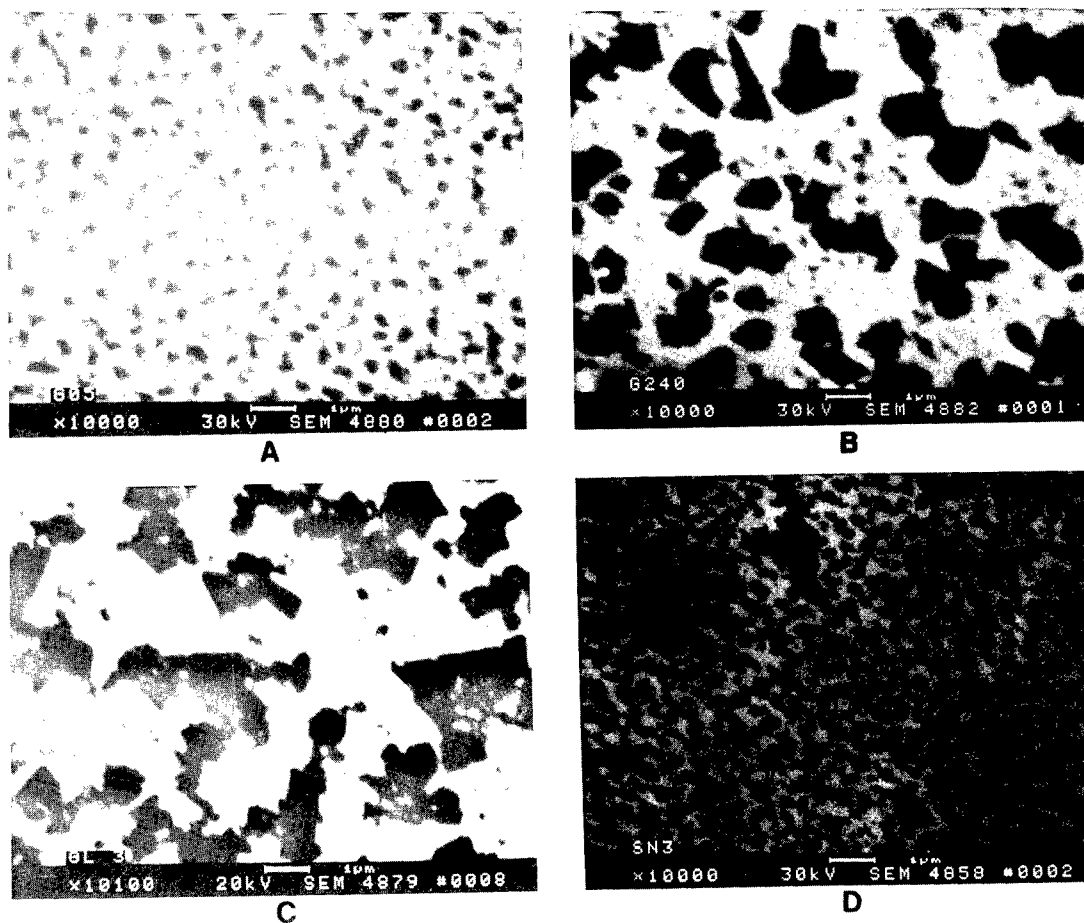e, developing Au rich and Ge rich regions. The As predominantly occupies the same regions as the Ge and the Ga shows a random pattern. The SIMS depth profiles in Figure 5 indicate that the 4 elements move rapidly during the 5 min. anneal (Figure 5a and 5b) and at a slower rate during the 3 hr. anneal (Figure 5c). As expected, less motion occurs with the presence of a Si_3N_4 cap (Figure 5d and 5e). The depth profiles indicate the presence of Ga and As at the outer surface, and a minimum before the metal-semiconductor interface. Since the levels of Ga and As were below the LOD for AES and much higher than the LOD for SIMS, we estimate the concentration of these species to be in a range of 0.0001 to 0.1% Ga and As at the top surface layer.

The backscattering spectra for a standard sample is shown in Figure 6a. The surface peak positions for Au and Ge are indicated. The diffusion of Au into the substrate and towards the surface is evidenced by the lowering and broadening of the Au peak. The spreading of the Ge peak to lower energies indicates diffusion of Ge into the Au layer. The shift of the leading edge of the GaAs substrate near 2.2 MeV to a smaller slope indicates the displacement of Ga and As from the substrate. Plateaus characteristic of compound formation after a 3 hr. anneal (Figure 6b) can be seen in the Au signal near 2.7 MeV and in the region around 2.3 MeV. The intermixing of Ge with either the GaAs or Au is apparent by the absence of the Ge peak after 3 hr. of annealing but the extent of Ge diffusion toward the substrate cannot be precisely determined. Very little change in the spectra was found between the as-deposited and 5 min. annealed capped samples. However, the interactions between the metals and the GaAs in the capped sample are nearly as complete as the uncapped sample after the 3 hr. anneal. This corresponds well with the Schottky barrier height measurements on Si_3N_4 capped diodes: the barrier heights approached the same value after 3 hr. of annealing.

The high resolution TEM micrograph in Figure 7 shows that the interface between the metal film and the semiconductor is very smooth with no spiking into the GaAs substrate. The GeO_x layer is approximately 2 nm. Figure 8 displays the TEM microstructural results. The as-deposited material (Figure 8a) is composed of randomly oriented grains with an average grain size of 30 nm. Microdiffraction and EDS show these grains to be pure Ge and pure Au. The sample annealed for 5 min. (Figure 8b) shows both the Au and Ge grains covering the entire thickness of the metal layer. The elongated Au grains (dark) average 400 nm in length

IMAGE FIELD = 25 μm

Ge

Ga

As

Au
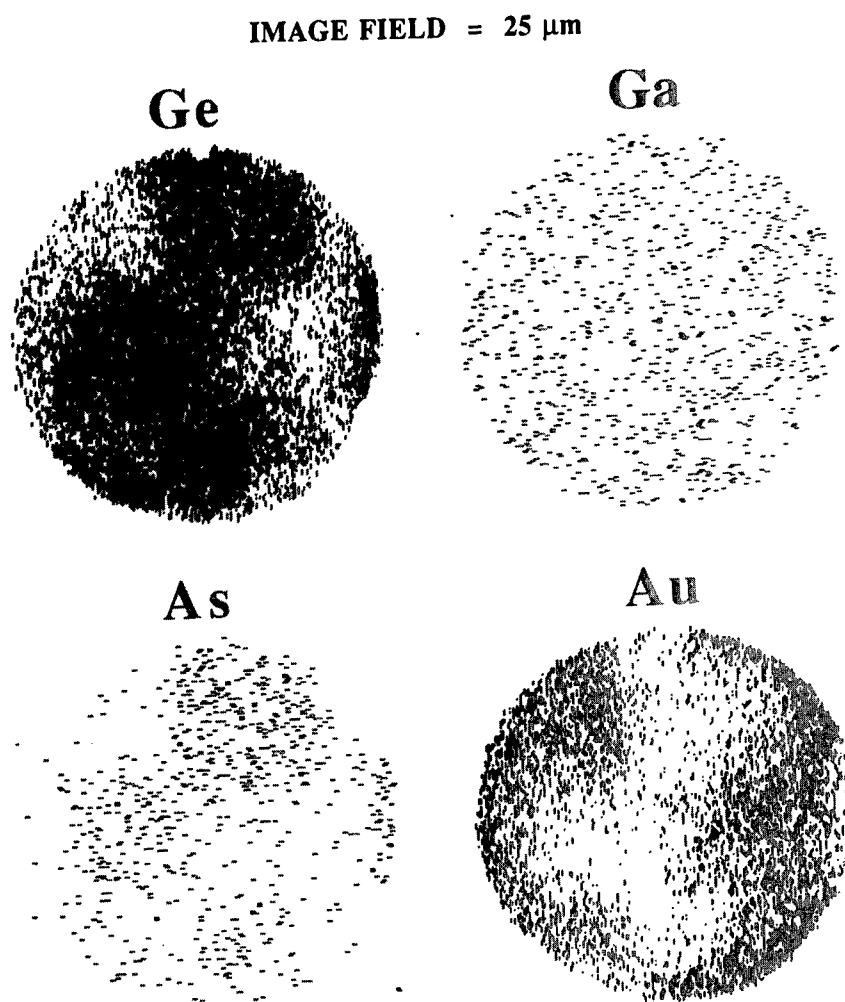
Figure 4. SIMS ion images of Au, Ge, As, and Ga distributions of a 3 hr. annealed sample.
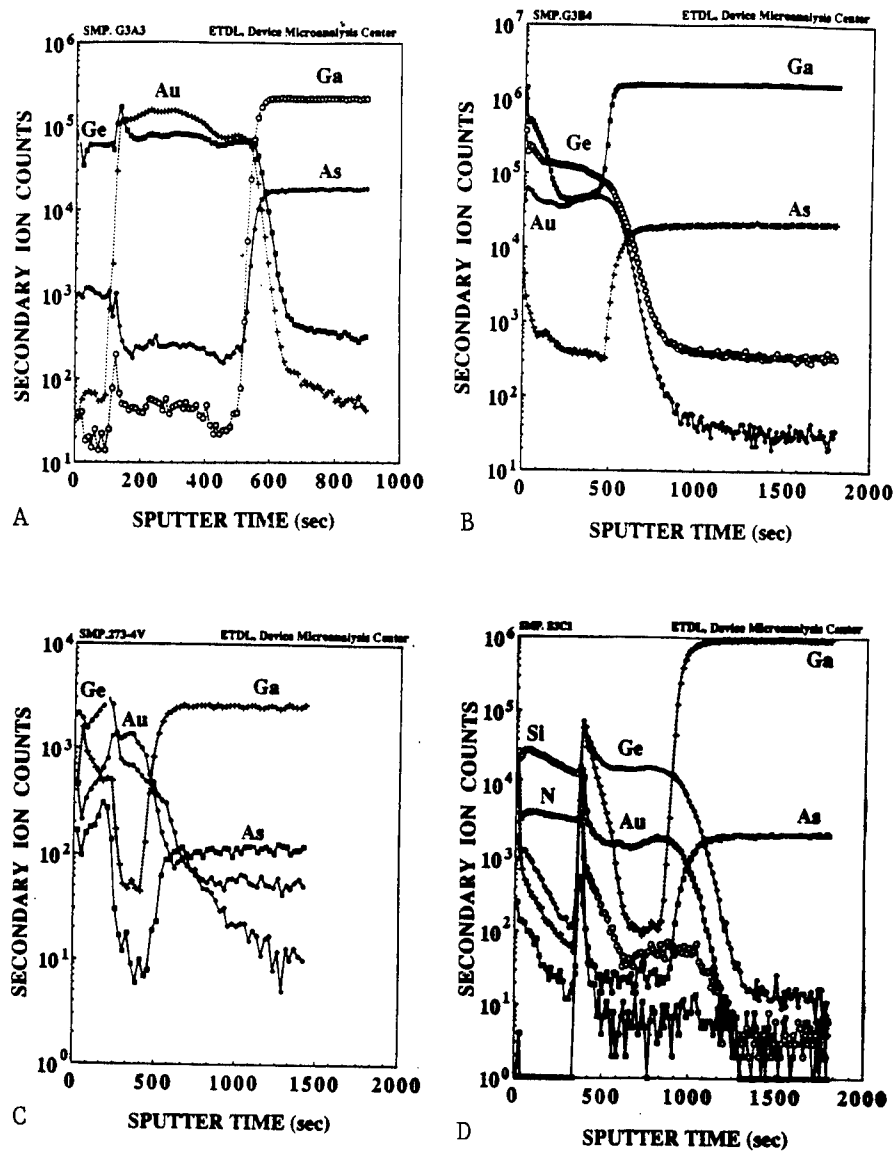
Figure 5. SIMS depth profiles for samples a.) as-deposited and b.) annealed for 5 min. for an uncapped sample, b.) uncapped and annealed for 3 hr., d.) annealed for 5 min. for a capped sample.
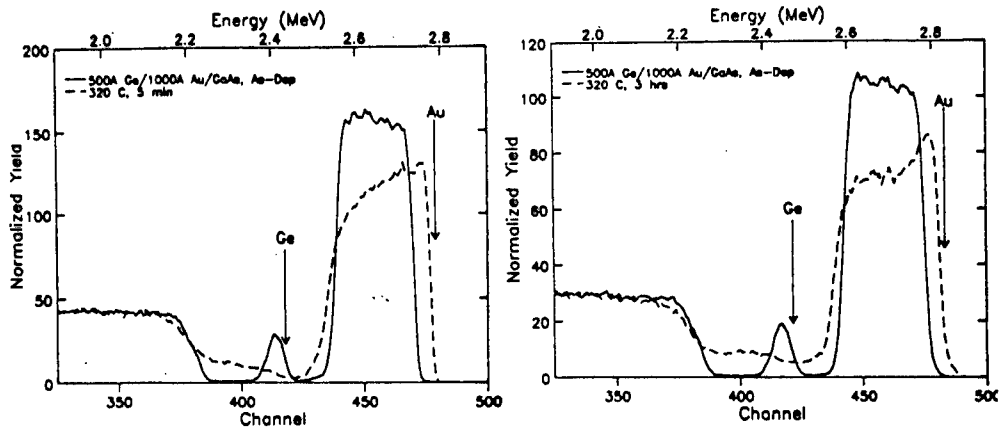
Figure 6. RBS spectra of a.) as-deposited and annealed for 5 min., and b.)as-deposited and annealed for 3 hr.
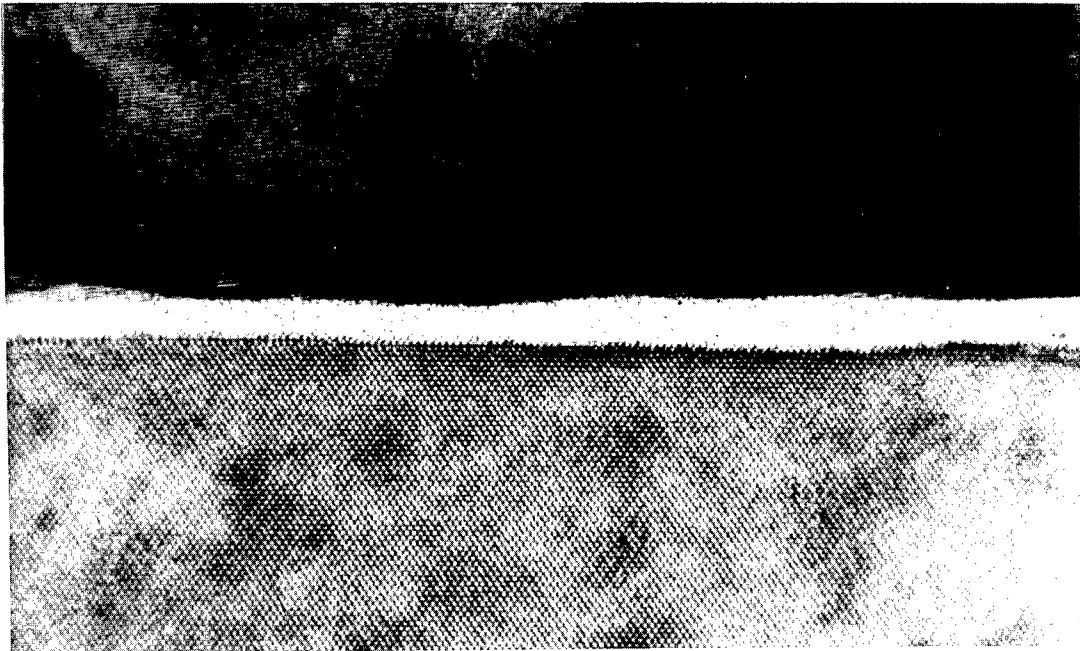
Figure 7. High Resolution TEM micrograph of smooth metal-semiconductor interface after 5 min. anneal. The oxide layer is approximately 2 nm.

while the Ge grains (light) average 250 nm in length. The Au grains exhibit a preferred orientation with the substrate, <111>Au // <100>GaAs. SAD and optical diffractogram analysis [23] identify an orthorhombic phase, with the approximate composition of Au_2Ga that appears after the 3 hr. anneal. The elongated Au_2Ga grains, approximately 1000 nm in length extend vertically from the top surface to the oxide layer (Figure 8c). These results do not rule out the existence of other Au-Ga or Ge-As phases. The capped sample, annealed for 5 min. (Figure 8d) is composed of only pure Au and pure Ge grains with the same orientational relationship as exhibited in the uncapped 5 min. anneal sample. However, the capped sample shows a less developed grain microstructure with respect to the uncapped sample.

DISCUSSION

The results support the hypothesis of Ge doping of the GaAs[4,15], but they indicate the mechanisms for ohmic contact formation is more complex than simply in-diffusion of Ge. Our results with the capped sample suggest that the Ge in-diffusion is enhanced by the out-diffusion of As, because r_c for the capped samples is a factor of 20 higher. The cap hinders As out-diffusion by blocking its evaporation [24]. Arsenic out-diffusion enhances Ge in-diffusion by increasing the rate of Ga out-diffusion. Ga vacancies are created which provide substitutional sites allowing the Ge to more readily diffuse into the GaAs. This explanation is based on the assumption that Ge occupies only group III sites. There has been some disagreement with this assessment and it has been suggested that the Ge occupies both group III and group V sites. Guha, et. al., [25] believe that the lattice distortion produced by the occupation of group V sites, alters the GaAs band structure to promote the formation of an ohmic contact.

The in-diffusion of the Ge on the group III sites must be assisted by the presence of Au since it is difficult or impossible to form an ohmic contact with the Ge rich samples. Au might enhance the in-diffusion of Ge by encouraging the out-diffusion of Ga and As by the reaction,

$$Au + GaAs \text{ --- } AuGa + \tfrac{1}{4} As_4$$
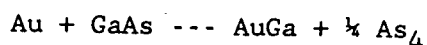
which has a more negative free energy of reaction than the decomposition of GaAs into Ga and As[21]. The process becomes more complex with the requirement of a Au-Ga phase. The Au-Ga phase does not form during the first minutes of annealing when substantial atomic movement occurs and elemental Ga and As appears on the Au surface. This rapid atomic movement, which includes the in-diffusion of Au during the relaxation of
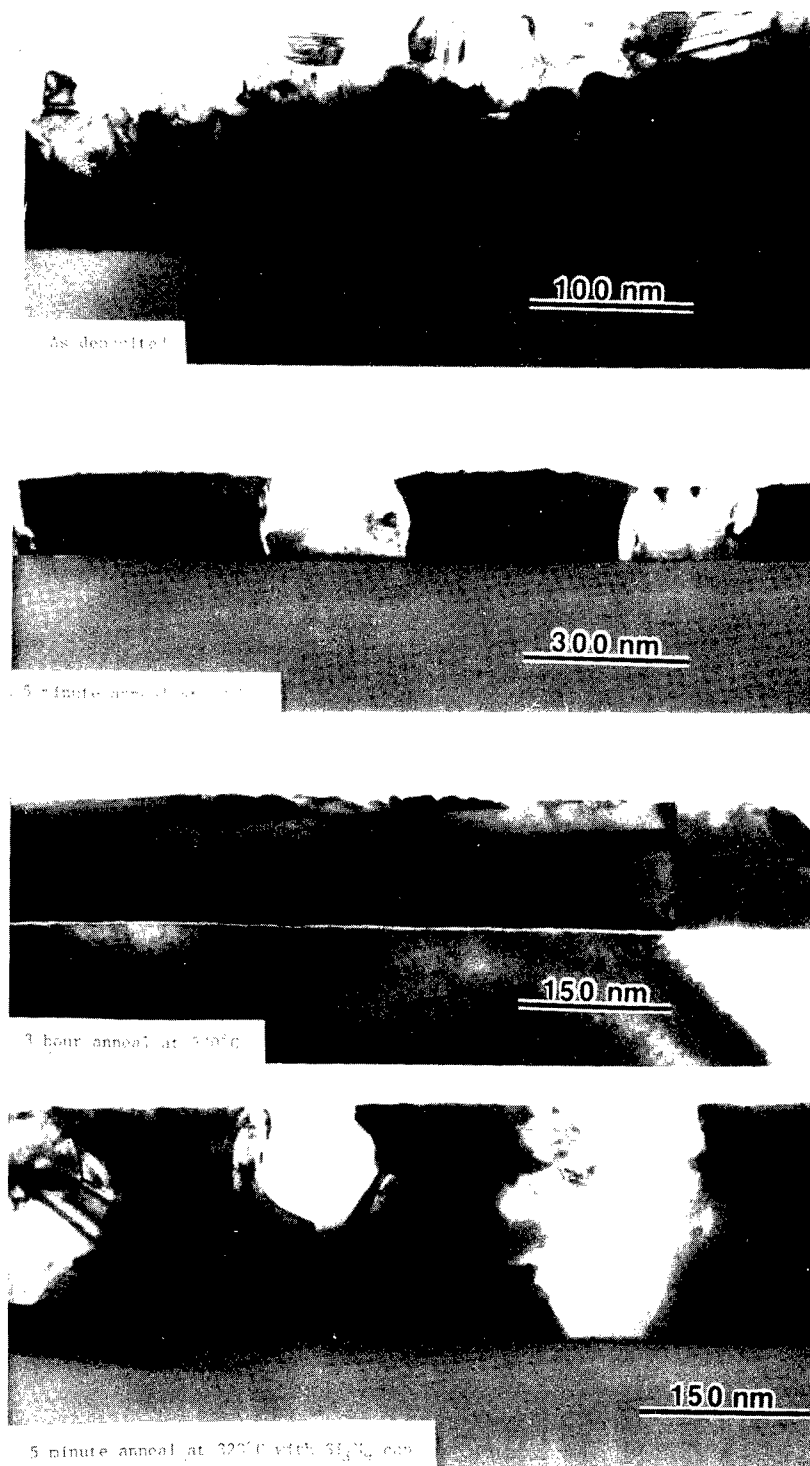
Figure 8. TEM micrographs of a.) an as deposited
sample, b.) a sample annealed for 5 min., c.) a sample
annealed for 3 hr., d.) a capped sample annealed
for 5 min.

the interface[22], probably is assisted by the creation of Au electronic states in the GaAs energy gap [2,26]. It is probable that the Au continues to promote the in-diffusion of Ge after the initial relaxation at the interface since the samples containing less Au form poorer contacts and require longer annealing times. The Au electronic states could independently promote the formation of an ohmic contact.

We agree with Iliadis[13] that the barrier height reduction, which promotes the formation of the ohmic contact, is primarily due to the interface induced chemical reactions and not to the formation of Au-Ga phases as has been proposed by others[17]. The barrier height reduction occurs in the first 5 min. of annealing but no Au-Ge phases are formed until after 3 hr. of annealing. Also the Au-Ga phase that forms during long anneals constitutes only a small portion of the interface. Our AES results support this by showing that the Au rich regions contain only a few percent Ga. Sinha and Poate[27] found that Au-GaAs Schottky diodes experience barrier height reduction when they are annealed, suggesting that interface mixing alone can account for a reduction in barrier height. This hypothesis becomes more credible by realizing that both the Au-Ge segregation and the reduction in barrier height occur during the first 5 min. of annealing.

Another hypothesis for the formation of ohmic contacts is a heavily As doped Ge phase in intimate contact with GaAs. We found that the Au and Ge segregate into a grain structure with the Ge grains extending the entire thickness of the metallized layer. It is possible that the As that is diffusing out of the GaAs could dope these Ge grains. A Si_3N_4 cap inhibits the out-diffusion of As[24] making more As available to dope the Ge, which should yield low resistance contacts. However, since the contacts actually have high contact resistances, we discount this hypothesis as a dominant mechanism.

TEM studies reveal the interface morphology to be smooth, differing from that of pure Au and alloyed contacts that often contain spiking of the metals into the semiconductor. The orientation relationship for the Au grains after 5 min. of annealing differs from that of pure Au[20].

CONCLUSIONS

Our results suggest that non-alloyed, Au-Ge ohmic contacts are formed by the combination of barrier height reduction and assisted in-diffusion of Ge following a 3 hr. anneal. The reduction in barrier height is caused by rapid intermixing at the interface during the first 5 min. of annealing and not by the formation of Au-Ga phases that form only after long anneals. Both the rapid mixing at the interface and the slow in-diffusion of the Ge are assisted by the out-diffusion of As and the in-diffusion of Au. These conclusions are supported by the high specific

contact resistances of tbhe capped and high concentration Ge samples. Our results do not support the theory of heavy doping of the Ge by As since samples with more As available had significantly higher specific contact resistances. The smooth metal-semiconductor interface is attributed to annealing substantially below the Au-Ge and Au-Ga eutectics. An orthorhombic Au_2Ga phase was found to exist after the 3 hr. anneal, but other Au-Ga and Ge-As phase possibilities cannot be ruled out.

ACKNOWLEDGMENTS

REFERENCES

1. J. S. Harris, Y. Nannichi, G. L. Pearson and G. F. Day, J. Appl. Phys. 40, 4575 (1969).

2. J. Gyulai, J. W. Mayer, V. Rodriguez, A. Y. C. Yu, and H. J. Gopen, J. Appl. Phys. 42, 3578 (1971).

3. C. Y. Chang, Y. K. Fang, and S. M. Sze, Solid- State Electron. 14, 541 (1971).

4. G. Y. Robinson, Solid-State Electron. 18, 331 (1975).

5. N. Yokoyama, S. Ohkawa, and H. Ishikawa, Jpn. J. Appl. Phys. 14, 1071 (1975).

6. H. Hartnagel, K. Tomizawa, L. H. Herron, and B. L. Weiss, Thin Solid Films 36, 393 (1976).

7. P. D. Vyas and B. L. Sharma, Thin Solid Films 51, L21 (1978).

8. O. Aina, S. W. Chiang, Y. S. Liu, and F. Bacon, J. Electrochem. Soc. 128, 2183 (1981).

9. M. I. Nathan and M. Heiblum, Solid-State Elect. 25, 1063 (1982).

10. S. Yasuami, Y. Saito, and A. Hojo, Jpn. J. Appl. Phys. 23, 379 (1984).

11. J. G. Werthen and D. R. Scifres, J. Appl. Phys. 52, 1127 (1981).

12. O. Aina, W. Katz, and B. J. Baliga, J. Appl. Phys. 53, 777 (1982).

13. A. Iliadis, J. Vac. Sci. & Technol. B5, 1340 (1987).

14. R. A. Stall, C. E. C. Wood, K. Board, and L. F. Eastman, Electron. Let. 15, 800 (1979).

15. W. J. Devlin, R. A. Stall , C. E. C. Wood, and L. F. Eastman, Solid-State Electron. 23, 823 (1980).

16. B. R. Pruniaux, J. Appl. Phys. 42, 3575 (1971).

17. S. Leung, T. Yoshie, C. L. Bauer, and A. G. Milnes, J. Electrochem. Soc. 132, 898 (1985).

18. N. Braslau, J. B. Gunn, and J. L. Staples, Solid-State Electron. 10, 381 (1967).

19. E. Kinsbron, P. K. Gallagher, and A. T. English, Solid-State Electron. 22, 517 (1979).

20. T. Yoshie, C. L. Bauer, and A. G. Milnes, Thin Solid Films 111, 149 (1984).

21. J. R. Lince, C. T. Tsai, and R. S. Williams, J. Mater. Res. 1, 537 (1986).

22. L. Brillson, R. S. Bauer, R. Z. Bachrach, and G. Hansson, Phys. Rev. B 23, 6204 (1981).

23. M. W. Cole, et al, to be published.

24. A. A. Lakhani, J. Appl. Phys., 59, 2082 (1986).

25. S. Guha, B. M. Arora, and U. P. Salvi, Solid-State Electron. 20, 431 (1975).

26. M. Jaros and H. L. Hartnagel, Solid-State Electron. 18, 1029 (1975).

27. A. K. Sinha and J. M. Poate, Appl. Phys. Lett. 23, 666 (1973).

Identification of a New Cyanide Countermeasure

Marland D. Dulaney, Jr.,Dr., Constaine Vlahacos,
Linda S. Pellicore, Dr., and John A. Wisler,Dr.
U.S.Army Medical Research Institute of Chemical Defense
Aberdeen Proving Ground, MD 21010-5425

DULANEY, VLAHOCOS, PELLICORE, WISLER

INTRODUCTION:

Hydrogen cyanide (HCN) is a rapidly acting poison which has recently been demonstrated to be a potentially dangerous chemical weapon (1). The presently accepted antidote is only for post-exposure use and consists of a methemoglobin former (sodium nitrite) and sodium thiosulfate as a sulfur source (2). This treatment regimen is effective, but sodium nitrite and sodium thiosulfate must be given intravenously, thus greatly limiting their use.

The development of a specific pretreatment compound which will not only protect against high doses (approximately 2 LD_{50}s) of cyanide but will also have no deleterious side effects has not been realized with present technology. However, if the concept of pretreatment were modified to include a compound or mixture which has minimal side effects and will protect the individual against cyanide long enough for a specific post-exposure antidote to be administered, then this goal is more likely to be achieved. The pretreatment should consist of a non-toxic cyanide binding agent and a sulfur source. The sulfur source needs to be absorbed well by liver cells so that it can easily reach the cyanide detoxifying enzyme rhodanese. The ideal pretreatment should also be effective orally.

One class of non-toxic compounds reported to possess anti-cyanide activity in vivo includes the alpha-ketocarboxylic acids, pyruvate (7) and alpha-ketoglutaric acid (AKG) (8). These compounds presumably act as cyanide "buffers" in that they tie up cyanide instead of hydrogen. Chemically they undergo nucleophilic attack by cyanide at the activated ketone moiety producing a cyanohydrin. Pyruvate and AKG are even more effective when administered concurrently with high doses of sodium thiosulfate (7,8).

Since the introduction of sodium thiosulfate as a cyanide antidote (3), considerable research effort has been directed toward the development of a better sulfur source. Thiosulfate is a highly charged molecule that does not easily penetrate into cellular spaces (4), and large doses of the drug must be administered to achieve a therapeutic effect.

The thiosulfonates (5) are some of the most active sulfur sources for cyanide detoxification, when tested in vitro, using the cyanide metabolizing enzyme rhodanese. These compounds are derivatives of thiosulfate. They differ from the parent compound in that they have only one negative charge rather than two. Their effectiveness in vivo as sulfur donors, however, is no better than that of thiosulfate unless given in conjunction with exogenous rhodanese (6).

Thus the rationale of this study is to characterize a cyanide antidote which utilizes the efficacy of the non-toxic cyanide binding agent AKG in conjunction with a thiosulfonate, which acts as a sulfur source. This should allow the thiosulfonate time to metabolize cyanide to thiocyanate. The protective effect of the combination of AKG and the thiosulfonate should be enhanced by the addition of a methemoglobin former such as hydroxylamine.

DULANEY, VLAHOCOS, PELLICORE, WISLER

The study described here is actually three distinct, interrelated phases. Each phase tests the conclusions of the other two phases using a different approach.

METHODS:

PHASE I: MOLECULAR MODELING

This phase involves the use of a computer model of the enzyme bovine liver rhodanese (EC 2.8.1.1). The covalent and tertiary structure has been previously described (9). The computer model was displayed and manipulated on an Evans and Sutherland PS390. The three-dimensional coordinates for bovine liver rhodanese were obtained from the Brookhaven Database. The three-dimensional coordinates for thiosulfate and the four thiosulfonates were generated by _ab initio_ calculations using Gausian 88.

PHASE II: IN VITRO RHODANESE STUDIES

Materials: Sodium 2-aminoethane thiosulfonate (2-AETS), sodium alanine thiosulfonate (ATS), sodium ethyl thiosulfonate (ETS) and sodium propyl thiosulfonate (PTS) (Fig. 1) were synthesized under contract by Dr. John Westley (University of Chicago). Bovine liver rhodanese was crystallized under contract by Dr. Marguerite Volini (University of Hawaii). All other chemicals were purchased from Sigma Chemical Co. (St. Louis, MO).
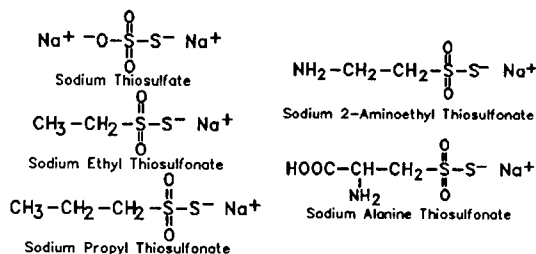
Figure 1. The structures of sodium thiosulfate and the sodium salts of the four thiosulfonates ethyl thiosulfonate (EST), propyl thiosulfonate (PTS), 2-aminoethyl thiosulfonate (2-AETS) and alanine thiosulfonate (ATS).

Preparation of Mouse Liver Mitochondria: The livers from male CD-1 mice were removed within one minute of sacrifice. The mitochondria were isolated using a sucrose/TRIS buffer (0.1 M). This method has been shown to be very efficient in isolating functional mitochondria (10). The resulting mitochondrial suspension was separated into two portions. One portion was immediately placed into a -70° C freezer for 24 hours. The remaining mitochondrial suspension was kept on ice and used within one hour.

DULANEY, VLAHOCOS, PELLICORE, WISLER

Rhodanese Measurements: Rhodanese activity of the fresh and frozen mouse liver mitochondrial suspensions, as well as the crystalline bovine enzyme, was followed by measuring thiocyanate production (2). The final concentrations of thiosulfate and the four alkyl thiosulfonates ranged from 0.01 to 25 mM for both the fresh and frozen mitochondrial suspension. The final protein concentration for the mitochondrial suspension ranged from 3 to 5 mg/ml. The effect of various concentrations of AKG (5 uM to 50 mM) on the sulfur donating activity of a fixed concentration of 2-AETS (2.5 mM) to rhodanese was also determined in both the fresh and frozen mitochondrial suspensions.

The crystalline bovine enzyme was stored at -70° C and was diluted in sodium phosphate buffer (0.1 M, pH 7.2) to a final concentration of 0.5 mg/ml immediately prior to use. The final concentrations of thiosulfate and the four alkyl thiosulfonates ranged from 0.01 to 5 mM.

Protein concentrations were determined using the Bradford method (11) with BSA as a standard. Rhodanese activity, indicated by thiocyanate production, was expressed as umoles of SCN⁻ produced per minute per milligram of protein.

PHASE III: IN VIVO STUDIES

Animal Husbandry: Male CD-1 mice (Mus musculus) from Charles River (25-30 g) were used in all experiments with ten animals in each group. The mice were quarantined on arrival and screened for evidence of disease before being released from quarantine. They were maintained under an AALAC accredited animal care and use program in plastic microisolator cages on hardwood contact bedding; they were provided commercial rodent ration and tap water ad libitum. Animal holding rooms were maintained at 21° C with 50% +/- 10% relative humidity using at least 10 air changes per hour of 100% conditioned fresh air. The mice were on a twelve-hour light/dark full spectrum lighting cycle with no twilight.

Lethality Studies (Pretreatment Only): The LD50 for sodium cyanide in mice of this weight range was found to be 6.2 mg/kg (i.p.). The control groups received either 5 mg/kg (i.p.) of thiosulfonate (2-AETS or PTS) or 250 mg/kg (i.p.) of AKG 15 minutes prior to injection of 6.7 mg/kg NaCN (i.p.). A third group received 2-AETS (5 mg/kg) and AKG (250 mg/kg), combined in the syringe for one minute, 15 minutes prior to cyanide. A fourth group received PTS (5 mg/kg) and AKG (250 mg/kg), combined in the syringe for one minute, 15 minutes prior to cyanide.

One group of animals received the combination of 2-AETS (50 mg/kg) and AKG (500 mg/kg) by oral gavage 15 minutes prior to NaCN (6.7 mg/kg, i.p.). The volume of dosing was 10 uL /10 g body weight.

All drugs were dissolved in normal saline with an injection volume of 100 uL/10 g body weight. The site of cyanide injection was always contralateral to the site of the pretreatment injection to reduce the chance of mixing of the compounds in the abdominal cavity.

DULANEY, VLAHOCOS, PELLICORE, WISLER

Cumulative mortality at one hour and time to death were recorded for each group. This study was conducted using mice, since they are very sensitive to cyanide, to determine if the combination of antidote compounds proposed would protect against the lethal effects of cyanide.

Cyanide Toxicity Scoring System: The duration and intensity of the toxic signs were monitored at one-minute intervals using a toxicity scoring system (12). The scale for monitoring and measuring the toxicity of injected cyanide is divided into four levels based upon the severity of the symptoms. Mice that either showed no signs of toxicity or had recovered, as demonstrated /by normal grooming and locomotor activity, received a score of zero. Animals that were in mild respiratory distress, used intercostal as well as diaphragmatic muscles, or exhibited a "splayed" hind foot, a common sign of cyanide toxicity, were assigned a score of one. Animals that had lost the righting reflex or were in moderate respiratory distress, used some intercostal muscles but mostly diaphragmatic muscles in breathing, were assigned a score of two. Animals that were in convulsions or in severe respiratory distress, characterized by audible "gasping" and only diaphragmatic muscle involvement, were assigned a score of three. Each mouse was scored individually at one-minute intervals for thirty minutes.

The scoring system described here was used to determine if the antidotes tested in these experiments would protect not only against the lethal effects of cyanide, but also against the incapacitation associated with this poison. For this reason, the animals were allowed to go through the entire spectrum of symptoms associated with cyanide toxicity with no steps taken to alleviate distress. It has been previously reported that death by cyanide, while aesthetically unpleasant, is painless and is an acceptable form of euthanasia (13).

Lethality Studies (Pretreatment:Treatment): Animals were pretreated with the combination of 2-AETS (50 mg/kg) and AKG (500 mg/kg) orally 15 minutes prior to NaCN (7.2 mg/kg, i.p.). One minute after cyanide exposure, the animals received hydroxylamine (5 mg/kg, i.m.) in the hind leg. Hydroxylamine controls were not pretreated and only received the post treatment.

Additional animals were pretreated with the combination of 2-AETS (5 mg/kg) and AKG (250 mg/kg) given i.p. 15 minutes prior NaCN (13.4 mg/kg, i.p.). One minute after cyanide these animals received another injection (i.m.) containing 2-AETS (5 mg/kg), AKG (250 mg/kg) and hydroxylamine HCl (5 mg/kg). In both of these pretreatment:treatment experiments only cumulative mortality was examined.

Data Analysis Plan: Differences in survival at one hour post cyanide exposure were analyzed for statistical significance by the Chi-square test. The data obtained from the toxicity scoring system were analyzed, when appropriate, using the Mann-Whitney U-test for non-parametric statistics. Comparisons against the cyanide control group became impossible after the eighth minute since all of the animals in this group had died. The data obtained from the

DULANEY, VLAHOCOS, PELLICORE, WISLER

rhodanese assay was analyzed for statistical significance using a two-way ANOVA followed by the Newman-Keuls test. In all three data analysis methods, a probability of $p < 0.05$ was considered significant.

RESULTS:

Crystalline Bovine Rhodanese: The sulfur donating activity of each of the four thiosulfonates was compared with sodium thiosulfate. The threshold for measurable sulfur donating activity, as followed by a one minute incubation period, was 0.1 mM for all four of the thiosulfonates and 0.5 mM for sodium thiosulfate (Fig. 2a). The rhodanese reaction was linear for concentrations up to 5 mM. At concentrations above 5 mM the reaction became difficult to follow due to the intense red color of the reaction product and the non-linearity of serial dilutions of the final product.

   All four of the thiosulfonates were significantly more effective sulfur sources to rhodanese than sodium thiosulfate. There was also a significant difference between the thiosulfonates which contained a positively charged nitrogen group in the side chain, 2-AETS and ATS, and the thiosulfonates, ETS and PTS, which did not have this group. It should be noted that in these experiments the two curves for 2-AETS and ATS were superimposed. (Fig. 2a).

Mitochondrial Rhodanese Activity: In the experiments using fresh mouse liver mitochondria there was a significant difference between the sulfur donating ability of all four of the thiosulfonates and thiosulfate. The threshold for measurable sulfur donating activity was 0.5 mM for 2-AETS, 1 mM for ATS, ETS and PTS, and 2.5 mM for thiosulfate. Among the four thiosulfonates, 2-AETS was a significantly better sulfur source than either PTS or ETS, which were both significantly better than ATS (Fig. 2b).

   A similar pattern was observed using freeze-fractured mouse liver mitochondria although, as expected, rhodanese activity was higher using this preparation. The threshold for enzymatic activity was found to be the same for all four of the thiosulfonates, 0.5 mM, with no change observed in the threshold for thiosulfate. 2-AETS was the best sulfur donor followed by PTS, ETS and ATS, respectively (Fig. 2c).

   AKG had no effect on sulfur donating activity of either 2-AETS or thiosulfate except at the highest concentration (50 mM) in both the fresh and freeze-fractured preparation. This concentration of AKG reduced the sulfur donating activity of 2-AETS by 30% using the fresh mitochondrial suspension and by 10% using the frozen enzyme suspension.

Lethality Studies (Pretreatment): Sodium cyanide alone (6.7 mg/kg, i.p.) resulted in 100% lethality within eight minutes. Pretreatment with 250 mg/kg of AKG alone significantly reduced ($p < 0.01$) the mortality of this dose of cyanide to 25%. Prior administration of either 2-AETS or PTS alone had no effect on mortality (Fig 3a).

DULANEY, VLAHOCOS, PELLICORE, WISLER
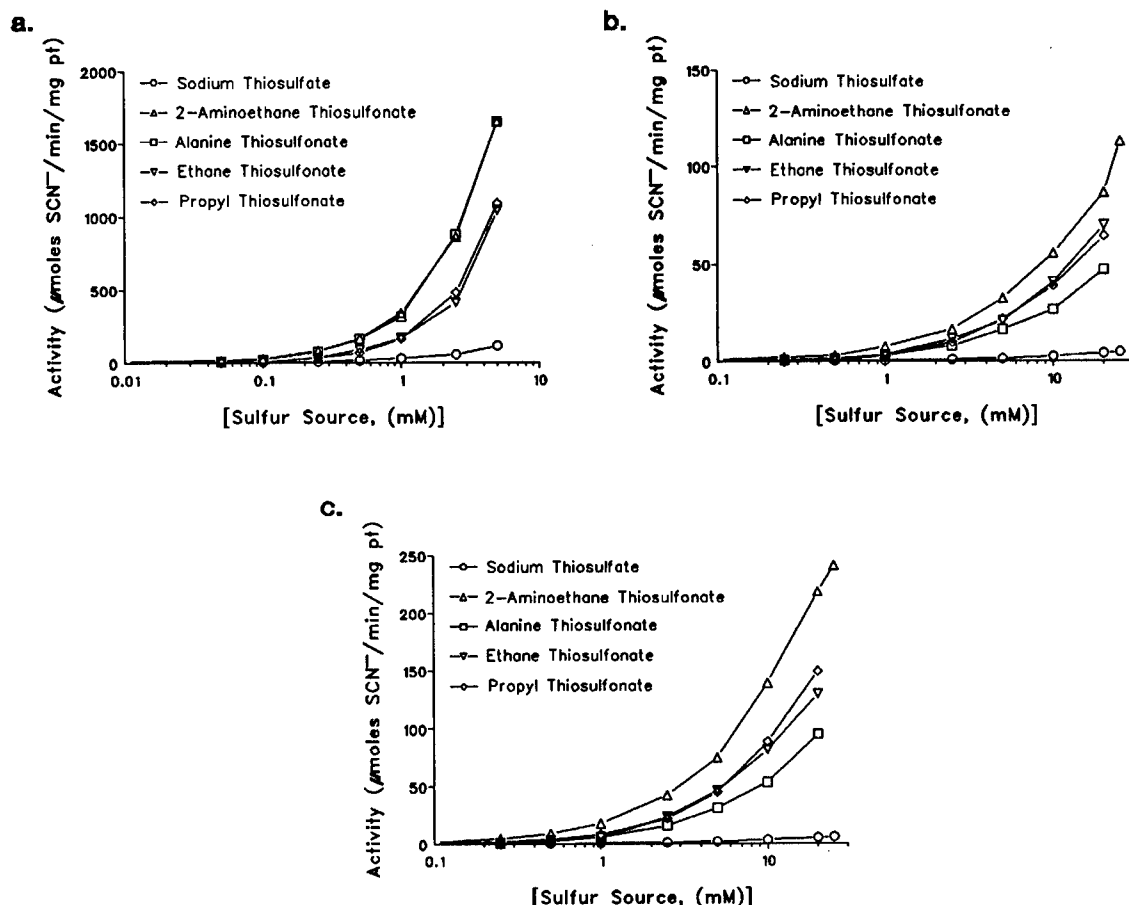
Figure 2: The sulfur donating activity of sodium thiosulfate and the four thiosulfonates using (a) crystalline bovine rhodanese, (b) fresh mouse liver mitochondria, or (c) freeze-fractured mouse liver mitochondria as the source of the enzyme.

When AKG and 2-AETS were combined in the same syringe prior to administration, the lethality of the injected cyanide was reduced to 10% (p <0.01). PTS actually antagonized the efficacy of AKG as a cyanide antagonist resulting in 90% lethality **(Fig 3b)**. AKG and 2-AETS combined and then given orally reduced the mortality of injected cyanide by 80% **(Fig. 4)**.

Toxicity Scores: The cyanide control animals reached the highest toxicity score by the second minute and continued in this stage until death **(Fig. 5b)**. Although a significant number of the animals pretreated with AKG alone survived, many of these animals also showed severe signs of cyanide poisoning. Only three of the twenty AKG-treated animals did not lose

DULANEY, VLAHOCOS, PELLICORE, WISLER

consciousness at some period during the experiment, and 25% of the animals showed convulsions for more than one minute during this period.
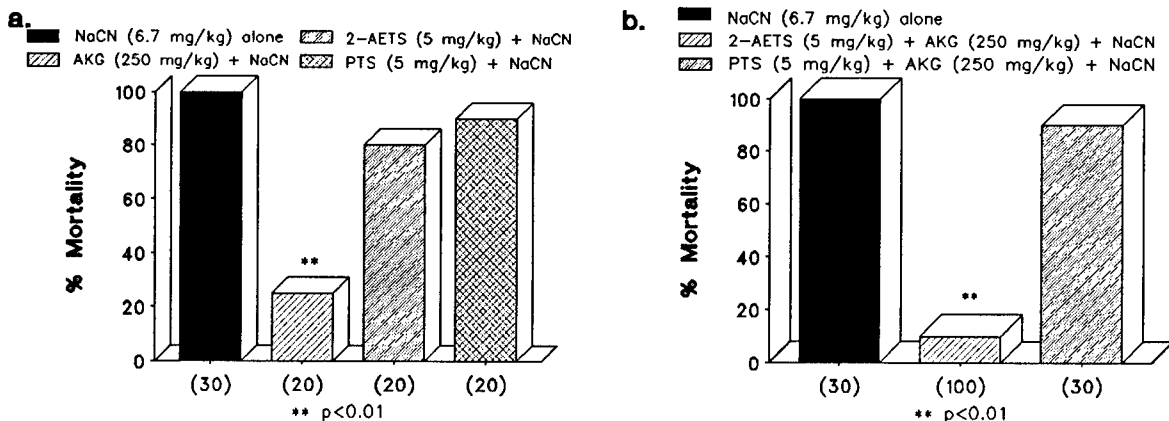
Figure 3: The effectiveness of various pretreatments alone (a) or in combination (b) administered 15 minutes prior to sodium cyanide (6.7 mg/kg, i.p.) on cumulative mortality in mice in the first hour. The numbers under each bar represent the number of animals in each group. The asterisks indicate statistical significance versus cyanide controls at the indicated probability level.
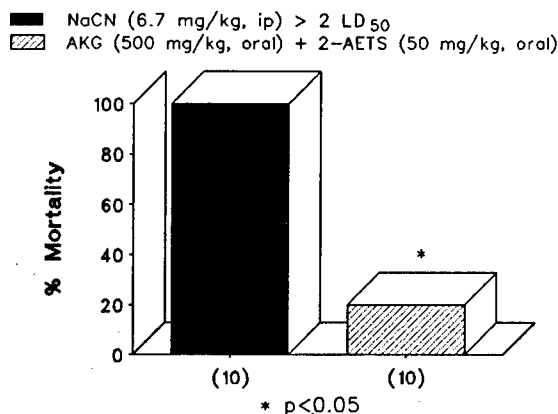
Figure 4: The oral effectiveness of the combination of AKG and 2-AETS 15 minutes prior to sodium cyanide (6.7 mg/kg, i.p.) on cumulative mortality in mice in the first hour. The numbers under each bar represent the number of animals in each group. The asterisks indicate statistical significance versus cyanide controls at the indicated probability level.

498

DULANEY, VLAHOCOS, PELLICORE, WISLER

All of the animals that were pretreated with either 2-AETS or PTS alone and survived the experiment showed significant convu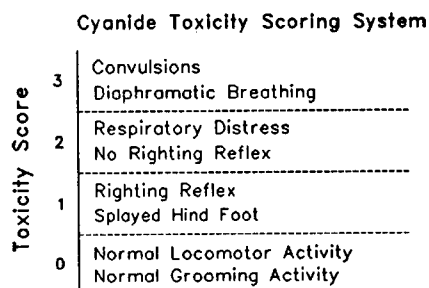lsions and/or diaphragmatic breathing during the majority of the experimental period (Fig. 5b). The animals that received the combination of AKG and PTS were also in severe respiratory distress through the first 15 minutes of the experimental period and all of them convulsed for at least one minute. However, of the animals which received the combination of 2-AETS and AKG only 22% ever lost consciousness and only 11% of these animals showed any signs of convulsion or heavy respiratory distress (Fig. 5b).

a.

Cyanide Toxicity Scoring System

Toxicity Score	
3	Convulsions Diaphramatic Breathing
2	Respiratory Distress No Righting Reflex
1	Righting Reflex Splayed Hind Foot
0	Normal Locomotor Activity Normal Grooming Activity

b. —•— NaCN alone (6.7 mg/kg) —○— PTS (5 mg/kg) + NaCN
—▽— AKG (250 mg/kg) + NaCN —◇— AKG + 2-AETS + NaCN
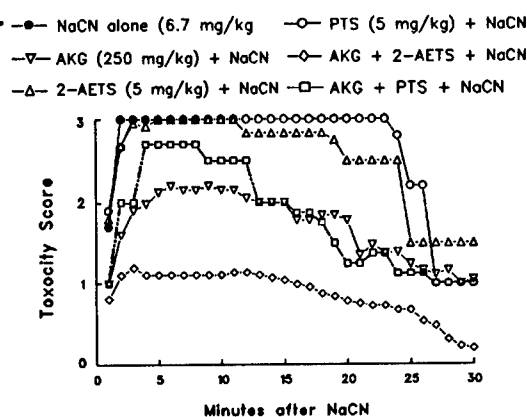—△— 2-AETS (5 mg/kg) + NaCN —□— AKG + PTS + NaCN

Figure 5: The left panel (a) shows the toxicity scoring system used in these experiments. The right panel (b) shows the toxicity scores for the animals in Figure 3. Note that there were no survivors in group 1 (NaCN alone) after 8 minutes. Except for cyanide controls, only animals which survived the entire experiment were used in the scoring.

Lethality Studies (Pretreatment:Treatment): Sodium cyanide at a dose of 7.2 mg/kg (i.p.) also resulted in 100% mortality. The mortality in animals which were pretreated orally with the combination of AKG and 2-AETS followed by the treatment, containing hydroxylamine, one minute after cyanide exposure was very significantly ($p < 0.01$) reduced to 6% (Fig. 6a). The single death occurred within thirty seconds which is indicative of an i.v. injection of cyanide. None of the surviving animals showed any moderate or severe signs of cyanide toxicity and they began normal grooming activity shortly after the treatment was given. Hydroxylamine alone reduced the mortality by 50%, however, all of the surviving animals showed severe symptoms of cyanide toxicity which continued for up to 50 minutes.

The animals exposed to the highest test dose of cyanide (13.4 mg/kg, i.p.) all died within one minute. None of the animals which received the pretreatment (i.p.) containing AKG and 2-AETS and the treatment containing 2-AETS, AKG and hydroxylamine showed any signs of cyanide toxicity except

DULANEY, VLAHOCOS, PELLICORE, WISLER

splaying of the hind foot in the minute between cyanide exposure and the
treatment. None of the animals ever lost consciousness and all were showing
normal activity almost immediately after the treatment (Fig. 6b).
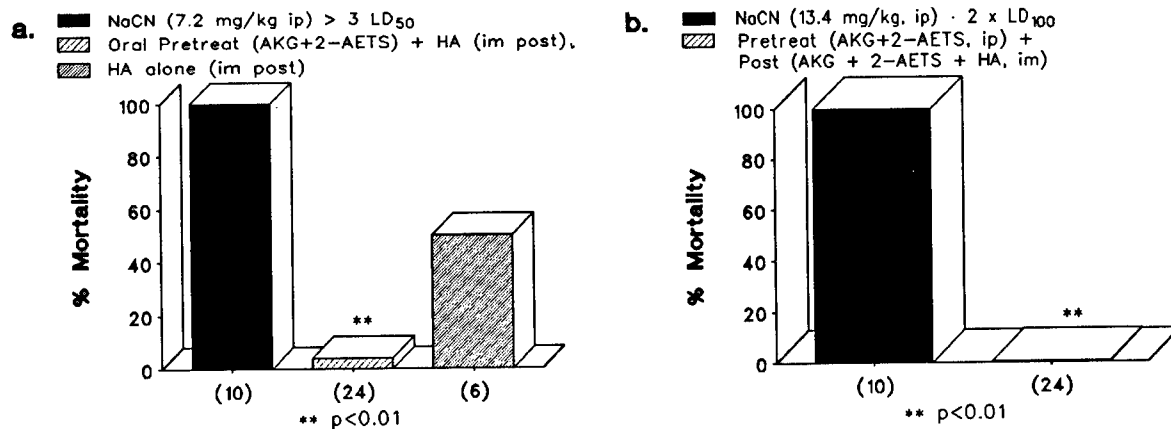
Figure 6: The effectiveness of oral pretreatment (a) with AKG and 2-AETS
administered 15 minutes prior to sodium cyanide (7.2 mg/kg, i.p.) followed
in one minute by hydroxylamine or hydroxylamine alone post-exposure on
cumulative mortality in mice in the first hour. The effectiveness of
pretreatment (b) with AKG and 2-AETS (i.p.) administered 15 minutes prior to
sodium cyanide (13.4 mg/kg, i.p.) followed in one minute by the combination
of AKG, 2-AETS and hydroxylamine on cumulative mortality in mice in the first
hour. The numbers under each bar represent the number of animals in each
group. The asterisks indicate statistical significance versus cyanide
controls at the indicated probability level.

DISCUSSION:

The data from these studies are evidence that the combination of AKG and
2-AETS is a very effective prophylactic antidote against cyanide when given
15 minutes prior to the cyanide challenge. There was significant protection
not only against the lethal effects of cyanide, but also against the
incapacitation associated with cyanide toxicity. This is an important point
since any pretreatment should protect against the unconsciousness and
convulsions normally associated with cyanide poisoning.

Some protection against the lethality was due to the presence of AKG (8);
however, the complete alleviation of the signs of toxicity seems to be the
result of combining the mixture of AKG and 2-AETS. This protective effect
was further enhanced by the addition of a methemoglobin former in the post
exposure antidote. The mechanism of action of this combination therapy has
yet to be firmly established. There are several interesting possibilities.

DULANEY, VLAHOCOS, PELLICORE, WISLER

The finding that AKG significantly reduces the lethality of injected cyanide confirms earlier reports on the efficacy of this non-toxic compound (8,14,15). This compound is also effective by either parenteral or oral routes of administration (14). It is interesting to note, however, that AKG did not significantly reduce the signs of cyanide toxicity, only the lethality. If the severity of toxicity is dependent upon the level of cyanide at some critical site(s) in the body, then AKG, acting as a cyanide "buffer," simply reduced cyanide levels sufficiently to allow for survival.

AKG may also be acting at another site, perhaps in the central nervous system, providing protection against the lethal effects of cyanide but not against the lesser signs of cyanide toxicity. This possibility has previously been raised (15).

Another interesting finding of these studies is the very significant increase in protection against high doses of cyanide resulting from the administration of a small amount of hydroxylamine. It is not surprising that some increase in protection against cyanide occurs from the concomitant use of a methemoglobin former. Methemoglobin also binds cyanide and this should increase the effectiveness of the pretreatment. What is surprising is that hydroxylamine acts very rapidly when used in conjunction with the pretreatment and that the protective effect is so marked using this low dose. Also this route of injection (i.m.) is favorable for the development of a treatment that can be self-administered. There are probably two reasons for this potentiation effect against cyanide. The first is easily apparent and embraces a simple approach to cyanide therapy. The other possibility is more controversial but may be equally important in the management of battlefield exposures to cyanide.

Hydroxylamine is much more effective in the combination therapy than alone because it is being administered in conjunction with another cyanide binding agent and a very potent sulfur source. This is the simplest explanation since it means that less methemoglobin formation is needed to assure survival. Also, there is less of a possibility for toxic side effects due to excess methemoglobin formation which has been a drawback in the development of this class of compound for either pretreatment or treatment (18).

The second possibility is that hydroxylamine is acting by another mechanism besides methemoglobin formation. Previous studies have noted that there is not a good correlation between methemoglobin formation by nitrites and protection against cyanide (19). Methemoglobin formers may act at some unknown site to block the lethal effects of cyanide. Hydroxylamine did contribute some protection against high doses of cyanide since some animals survived, however, the animals which survived exposure were incapacitated for almost one hour. In the context of this study it is only important that the addition of a small amount of hydroxylamine very significantly enhanced the effectiveness of the antidote.

These studies were designed to examine efficacy of two well known cyanide antagonists, AKG and hydroxylamine, when used in combination with 2-AETS. The effectiveness of these two compounds as cyanide antagonists has been well

DULANEY, VLAHOCOS, PELLICORE, WISLER

studied (8,18). However, the tremendous potentiation, especially against convulsions, that results from the addition of a small amount of 2-AETS is difficult to explain. It is clear that neither of the thiosulfonates showed any efficacy as a cyanide antidote alone. Thiosulfonates have been known to be excellent sulfur donors to rhodanese since the early 1950's (5); their efficacy in vivo, however, does not correlate well with their effectiveness as sulfur donors (6). To fully understand why the thiosulfonates, especially 2-AETS, are so effective, we have chosen to examine the three-dimensional configuration of rhodanese itself, especially the active site.

The computer model shows that this enzyme has one active site which is buried in a cleft between two relatively equal domains (Fig. 7a). The model used in this study agrees with previous authors that an important recognition site for substrates for rhodanese is a cationic site on the left side (Fig. 7a) of the cleft (9,16). Further analysis shows that a cluster of hydrophobic residues surrounding the right side (Fig. 7a) of the cleft may also be important (17).

Thiosulfonates do have structural characteristics similar to thiosulfate (Fig. 1) and they both almost certainly interact in a similar manner at the cationic site (Fig. 7b). However, these compounds also have unique characteristics which make them more effective sulfur donors to rhodanese.

The alkyl side chain reduces the net charge of the molecule from -2 to -1 making them less polar. This, plus the addition of an alkyl side chain, might make it easier for these compounds to cross biological membranes. Also, if the alkyl side chain itself had structural characteristics which could be recognized by transport systems in the cell membrane, i.e. protein carriers, then the compound might be able to enter the cell even more readily.

This suggestion is supported by the fresh mitochondria rhodanese studies (Fig. 2b) in which all of the thiosulfonates were more effective sulfur donors to rhodanese than was thiosulfate. The lesser negative charge should also reduce the electrostatic attraction of these compounds to the nitrogen atoms at the cationic site. This may allow these compounds to move closer to the active cysteine residue, where they could more easily donate their sulfur to form the persulfide.

However, if this was the only determinant of efficacy at the active site, then all of the thiosulfonate should be equally active regardless of the characteristics of the alkyl side chain. In both the crystalline bovine enzyme (Fig. 2a) and the frozen mouse mitochondria (Fig. 2a) studies the characteristics of the alkyl side chain significantly influenced the sulfur donating activity of the thiosulfonates. This suggests that another factor is involved in addition to attraction of the negatively charged portion of the compound to the cationic site. Based upon the bovine enzyme study, the second factor may be the result of electrostatic or stearic interactions with amine groups on the thiosulfonate side chain.

The model indicates that two regions could influence the activity of thiosulfonates at the active site. The first of these is the previously mentioned hydrophobic region. Highly polar compounds, like thiosulfonates, would be repelled by this region, forcing them closer to the active cysteine

DULANEY, VLAHOCOS, PELLICORE, WISLER

residue and increasing the probability that they would react with the active site. However, this type of interaction would be more non-specific in nature, acting more as a barrier against diffusion away from the active site than as a specific site which orientates the thiosulfonates to act as sulfur donors.
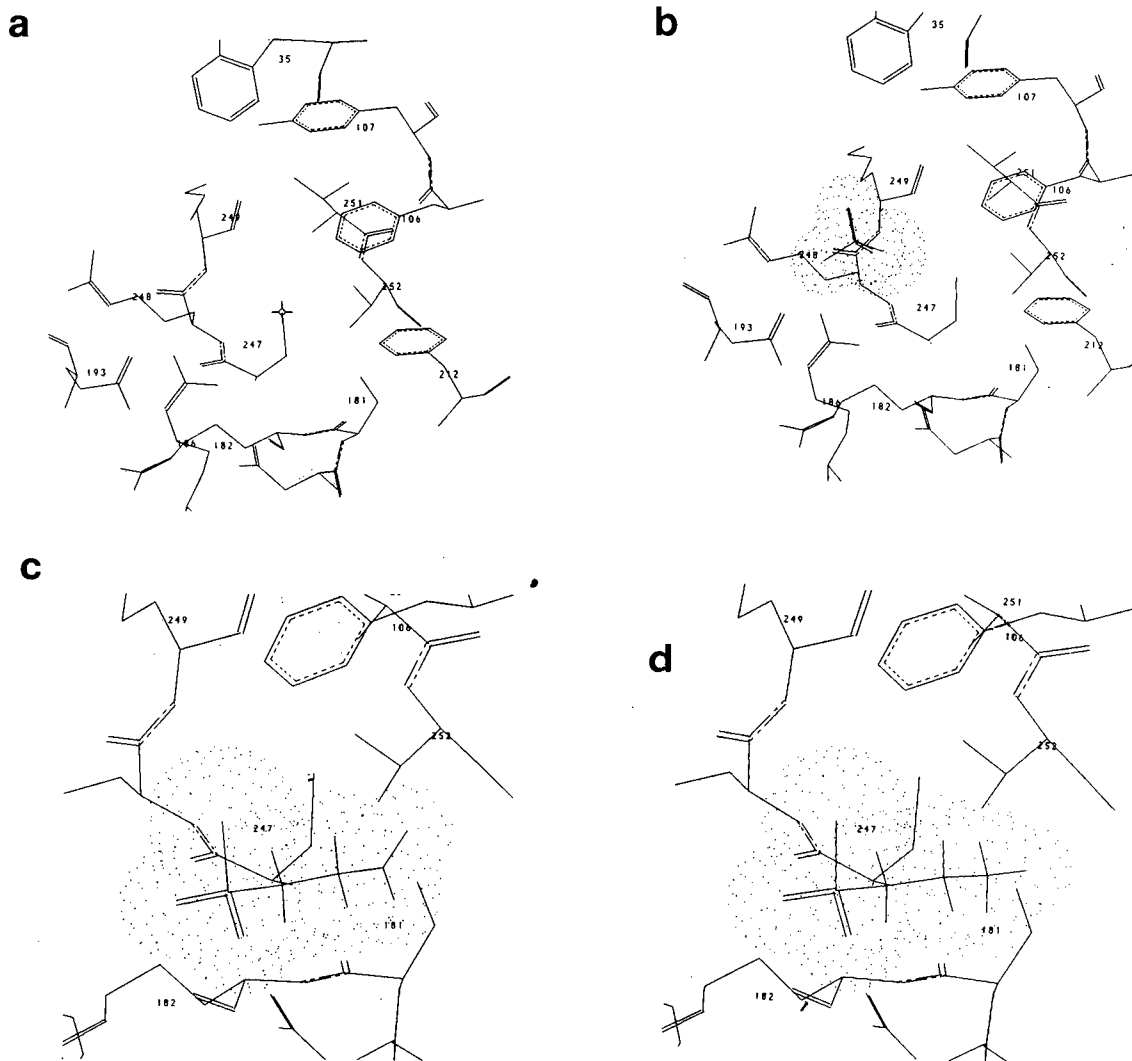
Figure 7: The molecular model of the enzyme crystalline bovine liver rhodanese. Panel a shows the active site of the enzyme. Panel b shows thiosulfate at the cationic site of the enzyme. The lower panels show the fit of 2-AETS (c) and PTS (d) at the pocket of oxygen atoms formed by SER 181 and THR 252.

DULANEY, VLAHOCOS, PELLICORE, WISLER

The second region is a group of hydrophilic residues immediately surrounding the active cysteine 247. This region is strongly implicated in hydrogen-bonding stabilization of the cysteine persulfide. Two of the residues in this region, serine 181 and threonine 252, contribute four oxygen atoms, two hydroxyl and two carbonyl, which form a pocket to one side of the active site. These electron-rich atoms, two of which may be involved in hydrogen bonding, could act to attract the positively charged amine groups of 2-AETS and ATS and orient them to donate sulfur to cysteine 247 (Fig 7c). This would not be true for PTS which also fits into the cleft but has no amine group to interact with the oxygen atoms (Fig. 7d). While this needs to be confirmed experimentally it does explain the effect of the amine groups on the activity of the thiosulfonates.

CONCLUSION:

These studies have found that a mixture containing the non-toxic cyanide binding agent, AKG, and a very potent thiosulfonate, 2-AETS, given as a pretreatment in combination with post-exposure treatment using hydroxylamine is a very effective countermeasure against at least 2 times the LD_{100} dose of cyanide in mice. The pretreatment is even effective orally and provides significant protection against the incapacitating effects of cyanide. The treatment is effective by intramuscular injection and can be given as long as one minute following cyanide exposure.

These studies have also made use of state-of-the-art molecular drug design technology to test new structures of sulfur donors for cyanide antidotes. The use of this technology, accompanied by laboratory studies run in parallel, should greatly increase the speed in which cyanide antidotes can be identified leading to eventual development, testing and fielding.

REFERENCES:

1. Heylin, M. ed. Chem. Eng. News. 66:23, 1988.
2. Westley, J. In: Biology of Cyanide. 61-76, 1981.
3. Lang, K. Biochem. Z. 259:243-265, 1932.
4. Gilman, A. et al., J. Biol. Chem. 147, 1946.
5. Sorbo, B.H. Acta. Chem. Scand. 7:32, 1953.
6. Frankenberg, L. Arch. Toxicol. 45:315-323, 1980.
7. Schwartz, C.,et al.,Tox. Appl. Pharmacol. 50:437, 1979.
8. Moore, S.J., et al., Tox. Appl. Pharmacol. 82:40, 1986.
9. Ploegman J. H., et al., Nature. 273:.24, 1978.
10. Toth, P.P., et al., Methods in Enzymology. 125:16-27, 1986.
11. Bradford, M. Anal Biochem. 72:248-254, 1976.
12. Hume A. et al., Human and Vet Toxicol. (accepted) 1990.
13. Sanford, J. Humane Destruction of Unwanted Animals. 21, 1976.
14. Dulaney, M.D., Jr., et al., The Pharmacol. 29(3):122, 1987.

DULANEY, VLAHOCOS, PELLICORE, WISLER

15. Yamamoto, H-a. Toxicol. Appl. Pharmacol. 99:415-420, 1989.
16. Mintel R. and Westley, J. J. Biol Chem. 241:3386, 1966.
17. Hol W.G.J. et al., Fund. Appl. Toxicol. 3:370, 1983.
18. Kruszyna R. et al., Arch Toxicol. 49:191, 1982.
19. Way, J.L. Annu. Rev. Pharmacol. Toxicol. 24:41, 1984.

Semiconductor Optical Waveguide Devices (U)

*Mitra Dutta, Dr.,
U.S. Army Electronics Technology & Devices Laboratory,
LABCOM, SLCET-ED,
Fort Monmouth, N.J. 07703-5000

INTRODUCTION

Photonics technology including sources, sensors, optical signal processing optical waveguides and interconnects and fiber optics will have strong impact on future defense systems. The greatest impact will be felt in the fields of communications, computing and sensors. Benefits of photonics include increased switching speeds, immunity to electromagnetic interference, massively parallel processing and interconnect functions and wideband information processing and storage. Enhanced performance of any of the elements of sources, switches, modulators, interconnects or sensors would increase the utility of optical techniques for signal processing applications. Further, as electronic systems become more complex and as the required speed of signal processing increases, wire interconnects become less effective. Integrated circuits are limited by pin and inter-connect counts, microwave interconnects are bulky and consume power and new system architectures require the distribution of many parallel high-data rate, large bandwidth signals. Among the system applications ham-pered by these interconnect considerations are supercomputers, phased array antennas, databases, local area networks and focal plane arrays. For a number of these applications optical interconnections may well be the answer, and optical waveguides and interconnects need to be developed in order to best take full advantage of the optical technology.

Heterostructures and multiple quantum wells in III-V semiconductors promise fresh approach and offer new options for high performance elec-tronic and optical devices, required for future high speed transfer, processing and control of information. Gallium arsenide (GaAs) based multiple quantum wells and heterostructures have opened up new areas of science and engineering and have excited a large amount of research. With molecular beam epitaxy (MBE) and metallo-organic chemical vapor deposition (MOCVD) techniques, it is now possible to prepare materials with a syn-thetic periodicity superimposed on the characteristic crystal lattice structure. The superlattice and multiple quantum well (MQW) materials have unique optical and electrical properties that are intrinsically different from the bulk compounds. Bandgap, lattice constants and refrac-

tive constants may be tailored in III-V heterolayers to confine carriers and radiation to impose built-in fields without the penalty of surface defects and to enhance mobilities, hot carrier effects, exciton absorption and optical nonlinearities. A number of devices not possible before in bulk semiconductors have been developed, making use of the new physics associated with these material and device structures. Efficient optical devices such as low threshold current quantum well lasers and quantum wire lasers, tailored bandgap infrared detectors based on intersubband absorption, modulators and switches making use of the quantum confined Stark effect and logic elements such as SEED devices are being researched and developed. In most of these above structures light propagates perpendicular to the layers, However when the light propagates parallel to the layers as in a waveguide additional new optical phenomena have been observed and makes possible some novel devices. All optical switching in waveguides can be particularly significant for applications in serial optical information distribution as optical pulses can be demultiplexed into either of two ports all-optically. The longer interaction length of the waveguide geometry allows more advantageous use of the optical nonlinearities.

The task of computing the exact mode shape and propagation constant of a waveguide can be extremely cumbersome in the case of a waveguide formed by numerous layers of differing index of refraction. The general procedure for solving for the guide is to assume a transverse mode shape of the electric and magnetic fields in each region (sinusoidal in guiding regions and exponentially decaying in non-guiding regions) and to match the tangential components of these fields at the boundaries (see Fig. 1). This is equivalent to matching the electric field and its first derivative at each boundary. The solutions are well known[2] for waveguides formed by a small number of layers with a high degree of symmetry. Solutions for guides formed by as many as seven layers have been reported[3]. However this method is quite time consuming when the number of layers increases further.

Kogelnik[1] has suggested a recursive algorithm using matrix representations of each layer which can compute the propagation constants of a waveguide formed by any number of layers. One application of such an algorithm is to compute the propagation constant and characteristic mode shape of the waveguides formed in multiple quantum well material which have numerous layers of differing composition and index of refraction. Approximations in which weighted averages of the indices of the constituent materials were used have been suggested[4]. However there is disagreement over which of the approximations provides better results.

<div align="center">THEORY</div>

Following the method of Kogelnik[1], a program was written to analyze waveguides formed by layers of materials with different indices of refraction (see Fig.2) in order to determine the propagation constant and mode
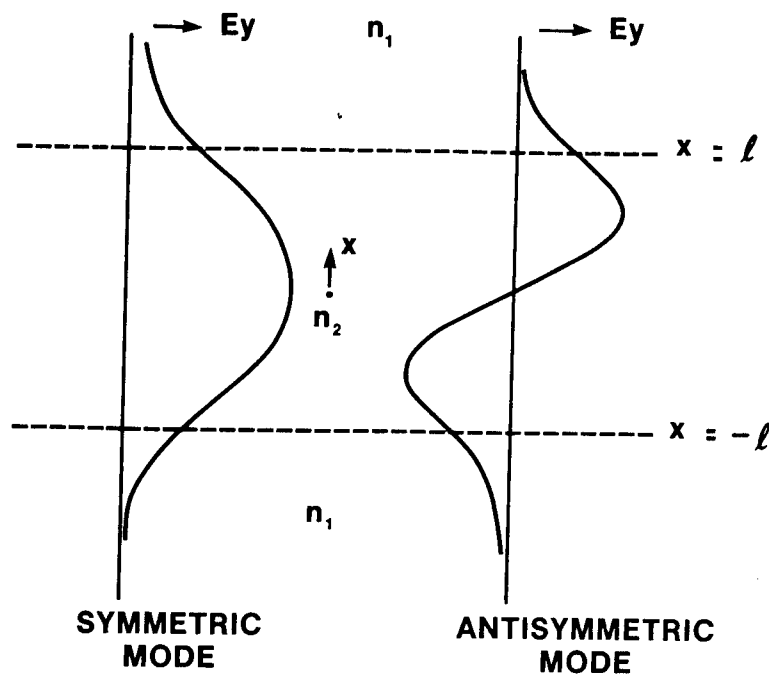
Fig. 1. Lowest order z-propagating symmetric and antisymmetric TE modes for symmetric slab waveguide of infinite extent in y.
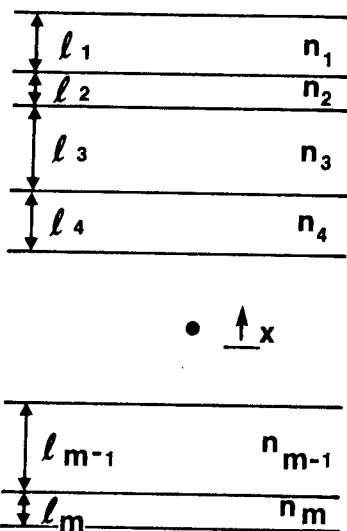

Fig. 2. Waveguide formed of different indices of refraction n_i and thicknesses l_i.

509

in each layer forming the waveguide. The effective index of the mode can then be computed as

$$n_{eff} = \sqrt{\frac{\beta}{k_0}} = \sqrt{\frac{\beta\lambda}{2\pi}}.$$

The electric field profile must satisfy the conditions that its value and first derivative be continuous at the boundaries between regions of differing refractive index. This is equivalent to requiring that the tangential electric and magnetic fields be continuous at these boundaries.

When the mode of the waveguide extends into a region of MQW material (which typically has many layers) the boundary conditions must be satisfied at each interface between layers of different refractive index to find the exact modes of the structure. In practice, approximate values are assumed and the "bulk" index of the region containing all of the layers is taken as single layer; its index is usually computed as a weighted average of the indices of the layers comprising the MQW region. Two of these averages are as follows

$$n' = \frac{n_1 l_1 + n_2 l_2}{l_1 + l_2},$$

and

$$n'' = \sqrt{\frac{n_1^2 l_1 + n_2^2 l_2}{l_1 + l_2}}.$$

In the following sections, results from the exact solutions will be compared with those obtained using the approximate formulas for n' and n".

WAVEGUIDE GEOMETRIES

A single symmetric slab waveguide was the first geometry to be analyzed. In this case the nonguiding regions were formed by AlGaAs with index of refraction n_1 and the guiding layer formed MQW material in which the layers were formed by alternating GaAs with index n_2 and AlGaAs (see Fig. 3). The index of refraction of the GaAs layers was taken to be 3.465 and that of the AlGaAs layer to be 3.315 for x = 0.3 (Ref. 5). The width of the guide normalized to the wavelength of the incident light was taken to be 1.

In order to accurately compare results in the three cases (n_{eff} (n_{exact}), n_{eff} (n'), and n_{eff} (n")), the MQW region is formed by an odd number of layers, with the first and last having index n_2 to differ from the surrounding region and to maintain symmetry. These layers are half the width of the other layers in the region in order to justify the use of

the formulas n' and n", which assume an equal number of layers of each type of material.
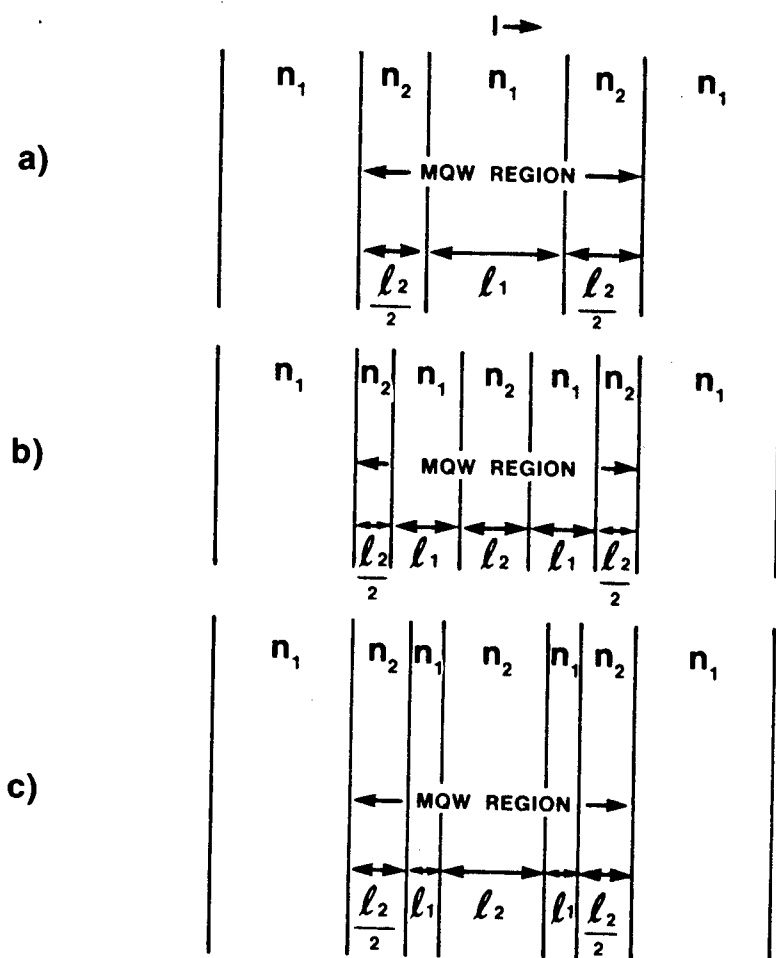
Fig. 3. Examples of symmetric slab waveguide geometries in which MQW layers form the guiding region. Calculations include changing the total number of layers with l_1/l_2 fixed, and changing l_1/l_2 with the total number of layers fixed. In all cases, the total thickness of the MQW region remains constant. (a). Five layer guide with $l_1/l_2 = 1$. (b). Nine layer guide with $l_1/l_2 = 1$. (c). Nine layer guide with $l_1/l_2 = 1/3$.

RESULTS

Using the single symmetric slab waveguide, the effective index of the guided modes of the guide was computed as a function of the number of layers in the quantum well region (see Fig. 3). In this analysis, the

511

ratio of layer thicknesses of the two materials forming the quantum well region and the total guide thickness were held constant. Three relative thicknesses of layers in the MQW region were examined; l_1/l_2 = 10, 1 and 0.1 where l_1 indicates the thickness of the AlGaAs layer and l_2 indicates the thickness of the GaAs layer. Two guided modes (the first symmetric and antisymmetric modes) are supported by the waveguide for values of layer thickness of 1 and 0.1, but only one is supported when the ratio is 10.
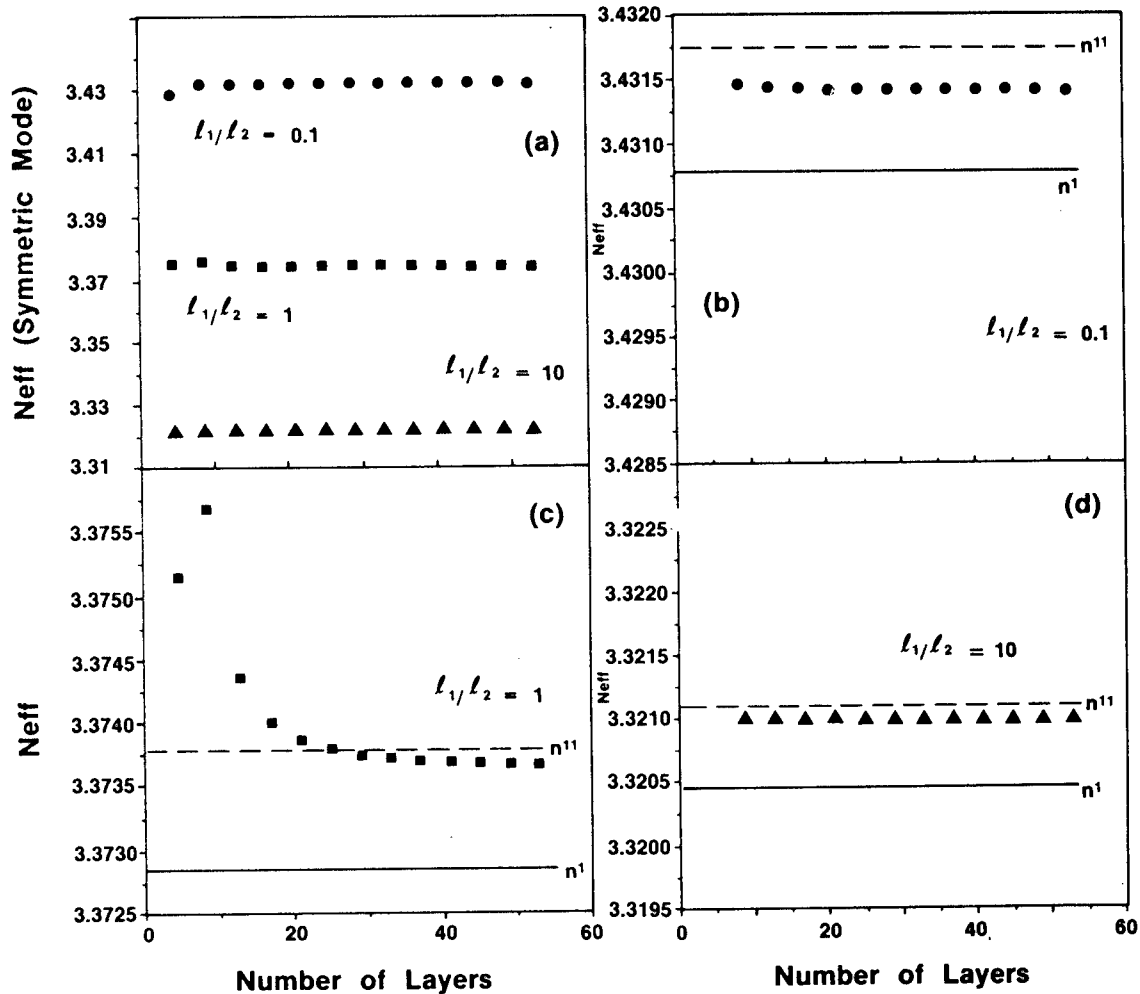

Fig. 4. Effective index of fixed slab MQW guide as a function of the number of layers. (a). Composite plot of ratios l_1/l_2 = 0.1, 1 and 10. (b). Enlargement of plot of l_1/l_2 = 0.1. (c). Enlargement of plot of l_1/l_2 = 10.

The variation in effective indices of the lowest order symmetric mode with the number of layers in the quantum well region are shown in Fig. 4.

Enlargements of these results for each of the three ratios of thicknesses of MQW layers are shown in Figs. 4b-4d. The number of layers was varied in increments of 4 (adding two periods of MQW material) from 5 to 53, corresponding to MQW guiding regions formed by 3 and 51 layers, respectively. The values of the indices are seen to approach a limiting value as the number of layers is increased. This was verified by computing the indices when the number of layers was 89; the variation in index for the cases of 53 layers and 89 layers was found to change by 0.006% or less in every case. The asymptotic values in each case are seen to be very close to the values obtained using either of the approximations using the average (indicated by the solid lines on Fig. 4b-4d) or weighted average index of the region (indicated by the dashed lines on Figs.4b-4d). These values are shown below in Table 1. The effective index computed exactly with 51 MQW layers is denoted $n_{eff}(53)$.

In a second set of calculations, the ratios of the thickness of the two different materials forming the region was varied while the total thickness of the 39 layer MQW region was held constant. The ratio of layer thicknesses l_1/l_2 was varied from 0.1 to 10. Again l_1 indicates the thickness of the AlGaAs layer with index n_1 and l_2 indicated the thickness of the GaAs layer with index n_2. The values of $n_{eff}(n')$ and $n_{eff}(n'')$ are compared with the values of n_{eff} for a region formed by 41 and 53 layer structures (see Fig. 5). The numerical values of the effective indices are shown below in Table 1 for values of $l_1/l_2 = 0.1$, 1, and 10.

	l_1/l_2	$n_{eff}(41)$	$n_{eff}(53)$	$n_{eff}(n')$	$n_{eff}(n'')$
Symmetric Mode	0.1	3.431406	3.431405	3.431079	3.431759
	1	3.373670	3.373649	3.372851	3.373778
	10	3.321042	3.321042	3.320451	3.321097
Antisymmetric Mode	0.1	3.374175	3.374180	3.373664	3.374490
	1	3.329382	3.329377	3.328877	3.329473
	10	MODE NOT SUPPORTED			

Table 1. Effective Index of Guided Modes for Ratios $l_1/l_2 = 0.1$, 1 and 10 Computed Exactly and Approximately Using n' and n".

DISCUSSION

As can be seen from the results above, both approximations for index of the MQW region result in values for the effective index of the mode which are in close agreement. Figure 4 and Table I show the effective index of the mode for the three different ratios of layer thicknesses in the MQW materials; when the number of layers is large enough that the change in index with increasing number of layers is negligible, the first
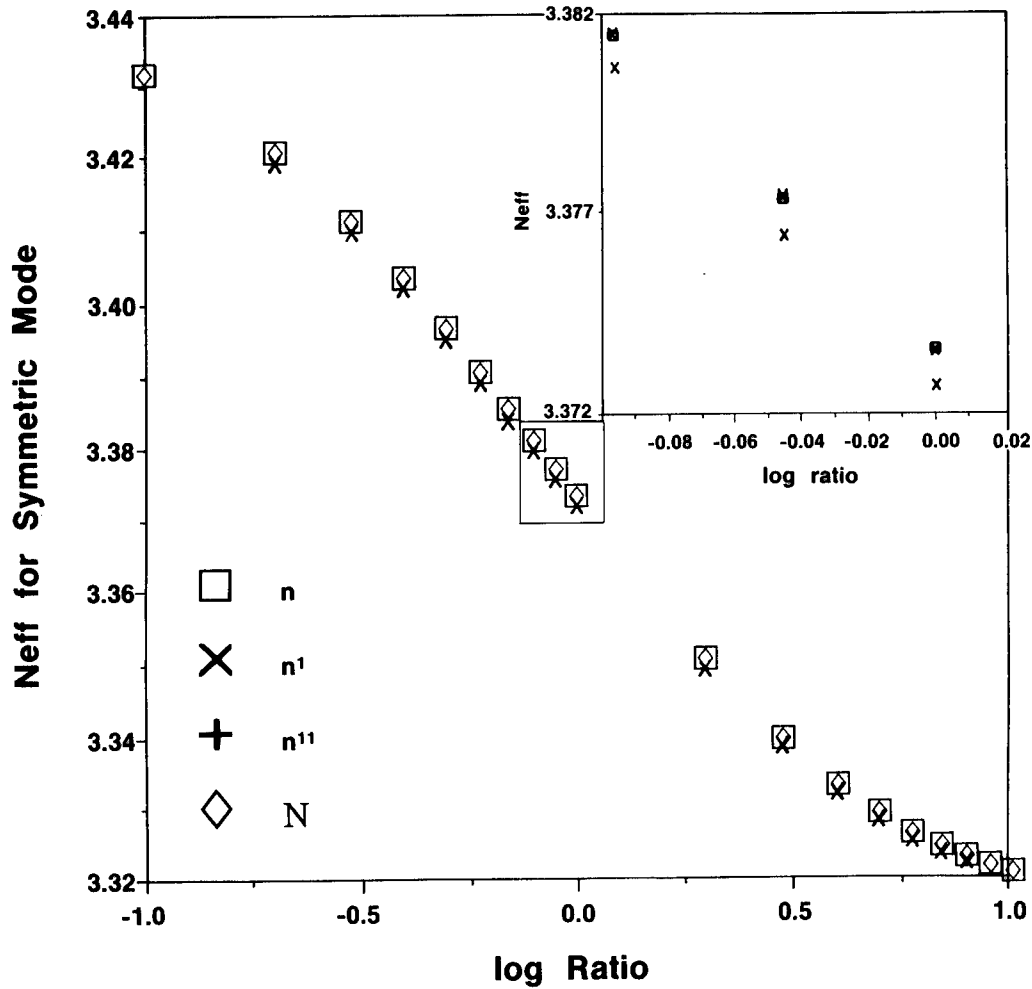
Fig. 5. Effective index of the symmetric mode of fixed width slab MQW guide as a function of the log of the ratio of layer widths. The four plots include effective index computed using (a) a 41 layer structure with 39 layers of alternating composition in the MQW region, (n); (b) a 53 layer structure with 51 layers of alternating composition in the MQW region, (N); (c) the average index n' of the region and (d) the weighted average index n" of the region.

approximation using n' underestimates the effective index of the mode while the second using n'' overestimates the value (compare $n_{eff}(53)$, $n_{eff}(n')$ and $n_{eff}(n'')$). For the symmetric guided mode of a large number of layers the average n' gives results which are in slightly closer agreement with the exact value found using the computer simulation when the ratio $l_1/l_2 = 0.1$ the results obtained are using the weight average n'' are closer to the exact values when $l_1/l_2 = 1$ and are nearly the same $l_1/l_2 = 10$. The larger relative changes in effective index for smaller numbers of layers is caused by the graininess of the refractive index within the layers. This is probably of little interest in practice as it is unlikely that MQW layers would be grown in these proportions and numbers. For the intermediate numbers of layers the exact solution crosses the approximations obtained using n'. This is of interest because in practical waveguide geometries, it is often the case that the waveguide structure would incorporate an intermediate number of layers. For example, in a waveguide which is $1\mu m$ wide formed by layers of thickness 60 Å and 440 Å, 20 periods of MQW's would be required.

Figure 5 and Table 1 show the values of the effective index as the ratio of the layer widths within the MQW material are changed while keeping the number of MQW's and the thicknesses of the entire region constant. The differences in effective index are so small that they are difficult to observe on the plots. In the insert the enlargement is shown. For 41 layer guides (formed by 39 layers of MQW) the approximation using n'' gives closer results than that using n'; the agreement is still closer when the 53 layer guide is used.

The second order mode was guided in geometries with the ratio of layer widths equal to 0.1 and 1. However no data is included for the antisymmetric mode because the index differences are even smaller. Because more of the antisymmetric mode resides in the region surrounding the guide when compared with to symmetric mode, it feels the effect of the MQW region less than the symmetric mode and the effective index depends less on the characteristic of the MQW region.

SUMMARY

The effective indices of the guided modes of slab waveguides formed using MQW materials are determined using a computer simulation to solve for the exact modes of the structures . These results are compared to those obtained using the two different formulas to approximate the index of the MQW region by using the average or a weighted average of the indices of the layers as the index of a single layer representing the entire MQW region in the slab waveguide. The effective indices of the modes found using either approximation are very close to the exact solutions. However, in most cases, the weighted average index of the region gives a slightly better value that the simple average.

ACKNOWLEDGEMENTS

REFERENCES

1. H. Kogelnik, in "Guided Wave Optoelectronics," T. Tamir (Ed.), Springer Series in Electronics and Photonics, 26, 7, (1988).

2. See for example, D. H. Lee, "Electromagnetic Principles of Integrated Optics," New York, John Wiley and Sons, 1986.

3. J. P. Donnelly, L. A. Molter and H. A. Haus, "Symmetrically Detuned Optical Three Guide Couplers in GaAs," presented at the Conference on Lasers abnd Electro-Optics (CLEO), Anaheim, CA, April 1988.

4. S. Ohke, T. Umeda and Y. Cho, "Optical Waveguides using GaAs-AlGaAs Multiple Quantum Well", Optics Communications, 56, 235, (1985).

5. S. Adachi, "GaAs, AlAs and AlGaAs: Material Parameters for use in Research and Device Applications," J. Appl. Phys. 58, R1, (1985).

Metabolic Resuce with a Precursor for Adenylate
Synthesis Stimulates Phosphate Utilization and
Prolongs Survival in Hemorrhagic Shock (U)

*Christopher P. Evans, CPT, MC
William P. Wiesmann, COL, MC
Walter Reed Army Institute of Research
Washington, D.C. 20307-5100

Hemorrhagic shock is a significant cause of morbidity and mortality in both the military[1] and civilian setting[2]. Despite prompt fluid resuscitation, detrimental metabolic effects due to decreased perfusion of vital organs persist and contribute to an unacceptably high mortality[3-9]. Data from the Vietnam War suggests that 10% of casualties died from exsanguination. Approximately 10% of soldiers that were saved by early resusitation still experienced a persistent shock syndrome and developed multiple organ failure that in most series carries a 60 to 90 % mortality. The lethal association of blood loss, shock, and organ failure is also apparent in civilian trauma where treatment delays are significantly less than those on a battlefield. Resuscitative efforts during the Vietnam era primarily focused on repletion of circulating blood volume with osmotic fluids and blood products. Few clinical means existed to improve the body's metabolic status, which deteriorates rapidly during the hemorrhagic shock state. This remains unchanged today.

Organ failure following shock is a complicated and sequentially amplified cascade that involves several events and mediators. The sequence of this cascade is organized as follows: 1. The loss of cellular energetics due to insufficient metabolic substrates and oxygen (reversible) 2. Subsequent cell swelling and leakage of precursors necessary for the resynthesis of high energy adenylates (reversible) and finally 3. The complete breakdown of cell integrity and mitochondrial function (irreversible). The key observation is that there are reversible components to this failure at the cellular and organ level which not only allow time for successful resusitation, but can be exploited as therapeutic targets. The challenge remains to develop an

adjunct therapy that will supplement metabolic needs during shock, improve organ circulation, and prevent tissue and capillary breakdown that ravages the trauma victim with infection and organ failure.

Adenosine triphosphate (ATP) is a critical purine nucleotide molecule that stores potential energy in phosphate bonds. Cleavage of these bonds releases energy that drives all functions necessary for cellular viability, including reactions that maintain cell membrane integrity. When ATP is depleted faster than it can be resynthesized or rephosphorylated, sodium enters the cell, potassium leaks out and the cell swells. Phosphorous accumulates following the breakdown of ATP, leaks out of the cell with other purines and enters the circulation. Thus, rapid increases in serum levels of potassium and phosphorous following shock are reflections of diminished cell function and energy production.

Liver function is crucial to survival during hemorrhagic shock. The liver generates most of the energy resource, glucose, required to maintain blood pressure[10]. Under aerobic conditions, glucose is converted to pyruvate during glycolysis, which occurs in the cytoplasm (Fig. 1). This yields two molecules of adenosine triphosphate (ATP). Pyruvate is transported into mitochondria and oxidized to acetyl coenzyme A. Here the citric acid cycle (Krebs cycle) and oxidative phosphorylation generate 34 more molecules of ATP. During anaerobic conditions that occur during shock, pyruvate is unable to enter the citric acid cycle and instead is converted to lactate. This metabolic end product accumulates in the blood as lactic acid.

Figure 1

CELL BIOENERGETICS
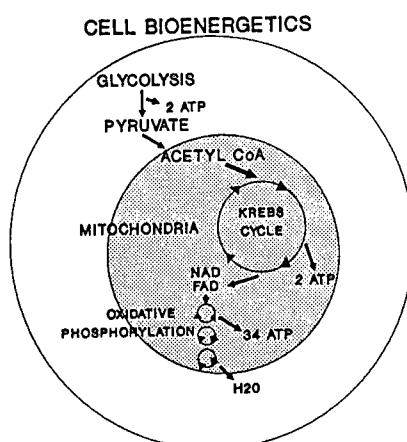
EVANS, WIESMANN

Other tissue beds are also ischemic in shock and generate lactate that worsens the problem. Under normal aerobic conditions the bloodstream carries lactate produced by glycolysis in skeletal muscle to the liver where it is converted to glucose via gluconeogenesis (Cori cycle, Fig.2). In the underperfused liver, however, gluconeogenesis is suppressed and vitally needed glucose is not regenerated. As an end result, ATP production decreases at a time when more energy is needed to protect against the insult. Compounding these metabolic deficiencies are decreased flow and perfusion. This results in lower delivery of substrates to tissues with a further worsening of energy production.

Figure 2

CORI CYCLE
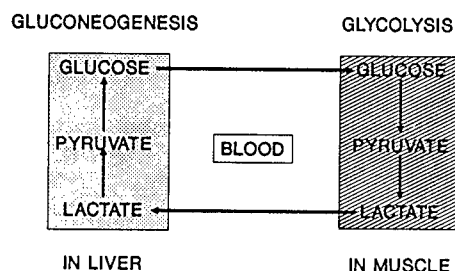
Many investigators have tried to intervene in this metabolic deterioration by providing substrates and energy sources such as glucose, ATP precursors or ATP itself. Most of these efforts have not shown a significant advantage over standard resusitative fluid therapy. Hypertonic glucose prolongs the decompensatory phase of hemorrhagic shock, thus improving survival[11]. Subsequent experiments, however, have shown that the extra glucose is not aerobically metabolized to pyruvate[12]. Apparently glucose infusions do induce vascular changes that cause transient increases in mean arterial pressure which probably account for its benefits.

Citric acid cycle intermediates such as fumarate have increased 24 hour survival in rabbits, but metabolic parameters were not made so the effect is unclear[13]. Circulatory changes may be occurring here as well.

Other investigations into the salvage of purine nucleotides focused on adenosine, a precursor of ATP[14]. Adenosine has a marked vasodilatory effect that supercedes its metabolic benefits. ATP given intravenously, chelates divalent cations producing various hemodynamic effects as well[15]. When coupled to magnesium chloride (ATP-MgCl$_2$), an

EVANS, WIESMANN

increased survival is observed in an animal model of shock[16-17]. The mechanism of action of ATP-MgCl$_2$ remains unclear. One possibility is that ATP-MgCl$_2$ improves organ perfusion through vasodilatation. Optimally, beneficial effects should be through energy generation and usage by the tissues. There is significant evidence that ATP-MgCl$_2$ does not cross cell membranes and is, therefore, unable to improve intracellular energetics[18].

We have studied 5-amino-4-imidazolecarboxyamide-riboside (AICA-riboside), a compound possessing the ability to improve ATP production through de novo purine synthesis (Fig. 3)[19-20]. AICA-riboside is a nontoxic, natural occurring substance that freely permeates cells. It does not cause vasodilatation[19]. AICA-riboside is phosphorylated to AICAR, which enters the de novo purine synthetic pathway midway and avoids earlier, highly regulated reactions. This phosphorylation step does not require 5-phosphoribosylpyrophosphate (PRPP) as a co-substrate, which is greatly decreased during ischemia. At low concentrations (<70uM), AICAR is converted to inosine monophosphate (IMP), then via adenylosuccinate (S-AMP) to adenosine monophosphate (AMP)[21]. This conversion to AMP requires the cellular investment of only one molecule of ATP and GTP (guanosine triphosphate). At high concentrations (>190um) IMP production is also increased, but nucleotide synthesis is reduced secondary to inhibition of the S-AMP lyase reaction. The amount of AICA-riboside infused during shock then, may have a significant influence on its effectiveness.

Figure 3

THE PURINE NUCLEOTIDE CYCLE
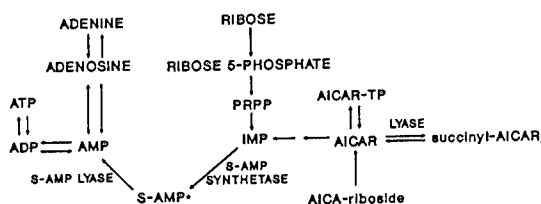
• S-AMP • adenylosuccinate

AICA-riboside infused intravenously in rats will increase IMP content 3-4 fold in cardiac muscle and 6-7 fold in skeletal muscle[19]. 25% of the infused AICA-riboside was metabolized to purine nucleotides, with a 10% increase in
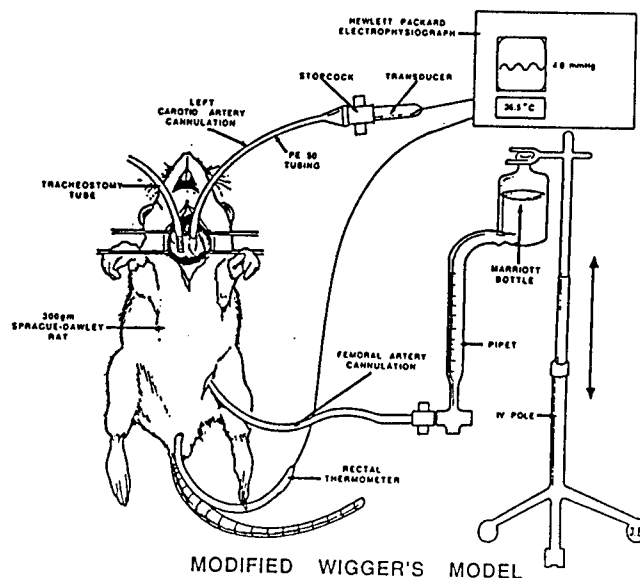
EVANS, WIESMANN

adenosine nucleotides in the myocardium. Mauser and others examined the effect of AICA-riboside on the rate of myocardial ATP synthesis during reperfusion after coronary artery occlusion in the dog[22]. After the occlusion phase, myocardial ATP levels decreased 50%. When the ischemic area was reperfused for 3 hours with AICA-riboside, a 9-fold increase in adenosine nucleotide·synthesis occurred. These studies have all examined short term effects of AICA-riboside in ischemic muscle and indicate that even small improvements in ATP synthetic ability could improve cellular energetics. There have been no previous studies on the effects of AICA-riboside in hemorrhagic shock and hepatic metabolism. Since AICA-riboside has few secondary effects on blood pressure, we hypothesized that its physiological and biochemical characteristics would be metabolically beneficial in an animal model of hemorrhagic shock.

METHODS:

We employed an isobaric (constant pressure) rat model of hemorrhagic shock to study the effects of AICA-riboside. In this modified Wiggers model, the animal is subjected to a variable stress - blood loss, to achieve a constant response - blood pressure[23]. Each animal's response to hemorrhage is individualized, with its blood pressure dictating the withdrawal or reinfusion of blood[24]. Following surgical preparation, (Fig. 4) blood is withdrawn until the mean arterial pressure is decreased to 40 mmHg.
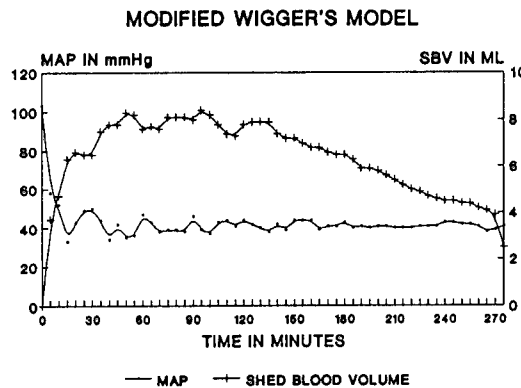
Figure 4

MODIFIED WIGGER'S MODEL

EVANS, WIESMANN

It is continually withdrawn or reinfused to maintain a blood
pressure of 40 mmHg (Fig. 5).

Figure 5

MODIFIED WIGGER'S MODEL

TIME IN MINUTES

— MAP —+— SHED BLOOD VOLUME

The removed blood is called the shed blood volume. Once the
animal reaches 40 mmHg, additional blood is withdrawn to
prevent the pressure from rising. This is the compensatory
phase, during which peripheral vascular resistance increases
in order to maintain central arterial pressure (Fig. 6).

Figure 6

STAGED MODIFIED WIGGER'S MODEL

COMPENSATORY PHASE
- INCREASED PVR
- INTERSTIAL FLUID MOVES
  INTO VASCULAR SPACE

DECOMPENSATORY PHASE
- DECREASE PVR
- TISSUE ISCHEMIA
- DECREASED ENERGY AVAILABLE
- DECREASED CIRCULATING BLOOD VOLUME
  DESPITE REINFUSION
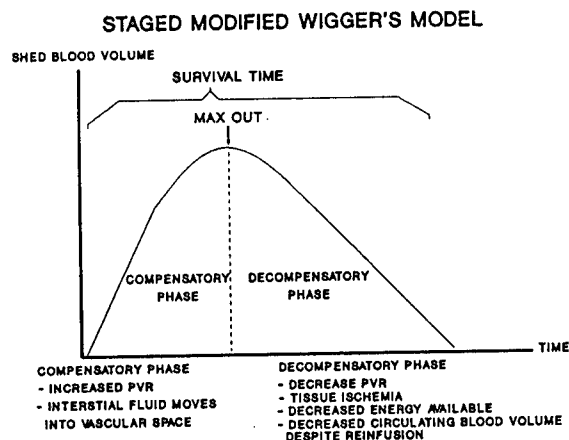
Blood flow to "less essential" organs such as skin, gut and
kidneys will be decreased in order to preserve perfusion to
the brain, heart and lungs. Also, interstitial fluid which
bathes the tissues will be drawn into the vascular space to
expand blood volume. These compensatory mechanisms will
eventually be exhausted at a point denoted as the maximum
output, "Max Out". From this time on, blood is reinfused in
order to maintain blood pressure at 40 mmHg. Peripheral
vascular resistance decreases as tissue ischemia worsens.

EVANS, WIESMANN

Despite fluid resuscitation, however, plasma leaks back out of the vascular space and blood pressure plummets followed by death. This is the decompensatory phase. Together the two phases combine to make the survival time.

The liver will become ischemic when 75% of "Max Out" is achieved. At this point hepatic blood flow is 30-50% of normal. Liver ATP is exhausted by "Max Out", with no further decrease during decompensation[4]. Glycogen breakdown results in a glucose peak in the early compensatory period, with levels dropping to near zero by time of death. If blood is reinfused early in the decompensatory phase, all animals will survive. After this, 75% will die despite volume resuscitation.

Experiments in our laboratory using this model have demonstrated that phosphorous is an early and sensitive marker of metabolic damage. In hemorrhaged animals, 66% of the total increase in serum phosphorous will occur by the end of the compensatory phase (30 minutes). If rats are only shocked for 6 minutes then resuscitated, phosphorous will briefly stabilize then continue to climb until a 57% increase is reached. Thus, despite restoration of blood volume, a lethal cascade has been activated. The effects of AICA-riboside, its inactive non-ribosylated form AICA, D-ribose and L-ribose were studied in this model. All compounds were prepared in lactated ringers solution at a concentration of 15 umoles, and injected in the jugular vein at "Max Out" and every 30 minutes thereafter.

Research was conducted in compliance with the Animal Welfare Act and other Federal statutes and regulations relating to animals. Experiments involving animals adhere to principles stated in the Guide for the Care and Use of Laboratory Animals, NIH publication 85-23, 1985.

RESULTS:
Shed blood used for fluid resuscitation during the decompensatory phase was analyzed for pH, electrolytes and phosphorous, and found to be statistically equal to in vivo baseline measurements.

Figure 7 compares survival times for animals treated with 15 umoles of AICA-riboside, AICA, D-ribose, or L-ribose at "Max Out". The compensatory phases of each group (which is prior to drug administration), were statistically equal. Decompensatory phases, and as a result survival times differed. Survival time in AICA-riboside treated animals was 2.6 fold longer than AICA treated rats (p<.005), 2.1 fold longer than the D-ribose group (p<.07), and 2.9 times longer

EVANS, WIESMANN

than rats receiving L-ribose (p<.01). The longer survival time in the D-ribose group as compared to the L-ribose group was not statistically significant (p=.5).

Figure 7

HEMORRHAGIC SHOCK: SURVIVAL TIME
AFTER METABOLIC INTERVENTION
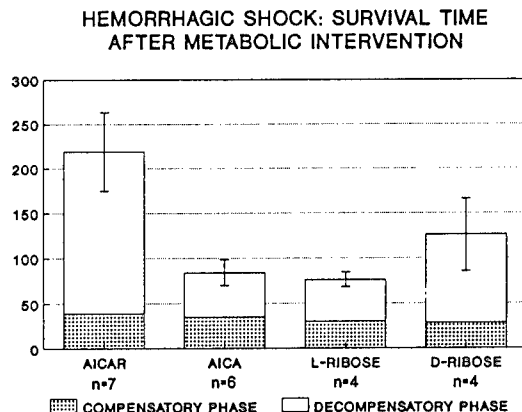
When AICA-riboside was administered approximately 10 minutes into the compensatory phase (when mean arterial blood pressure reached 40 mmHg), animals compensated 1.7 times longer than controls. This is shown in Table 1. The AICA-riboside treated rats took 111 ± 15 minutes to reach "Max Out", as compared with 67 ± 10 minutes for controls (p<.01). The AICA-riboside treated group also achieved a higher percentage of their predicted compensatory maximum output of shed blood volume than the control group: 95.6 ± 6% v. 82.8 ± 5%, p<.04 (studies in Sprague-Dawley rats have shown that predicted maximum shed blood mobilized to compensate for the hemorrhage is equal to 2.9% X body weight in gms).

Table 1
### EFFECTS OF AICAR ON
COMPENSATORY RESPONSE

|	Compensatory Phase in Minutes	Percent Max Out Attained
Group		
AICAR n=7	111.6 ± 15	95.7 ± 6
CONTROL n=7	67.1 ± 10 p<.01	82.9 ± 5 p<.04

EVANS, WIESMANN

Serum glucose levels during the hemorrhage protocol are shown in Figure 8. Equal baseline levels in AICA-riboside and control groups are followed by a higher peak at 30 and 60 minutes in the control group. From 90 to 240 minutes, AICA-riboside treated rats maintained higher levels, significant at 180 minutes (206 ± 68 mg/dl v. 127 ± 27 mg/dl, p<.07), 210 minutes (70 ± 28 mg/dl v. 21 ± 4 mg/dl, p, .03), and 240 minutes (47 ± 7 mg/dl v. 19 ± 4 md/dl, p<.02).

Figure 8

EFFECT OF AICAR ON
SERUM GLUCOSE

GLUCOSE IN MG/dL

TIME IN MINUTES

— AICAR n=7   —⊙— CONTROL n=7
AT 240 MINUTES, p<.02
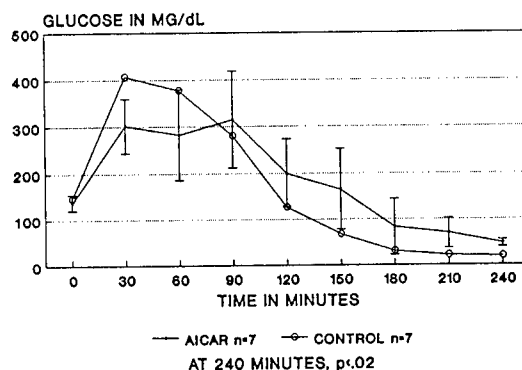
AICA-riboside and AICA serum inorganic phosphorous levels over a 4 hour shock period are shown in Figure 9. Phosphorous levels pre-treatment did not differ (AICA-riboside: 4.65 ± 0.48 mg/dl, AICA: 4.43 ± 0.46 mg/dl). After 240 minutes, however, the serum phosphate measurement of the AICA-riboside group increased less (0.93 ± 0.59 mg/dl) as compared to the AICA treated animals (3.1 ± 0.53 mg/dl). This was significant at p<.02.

Figure 9

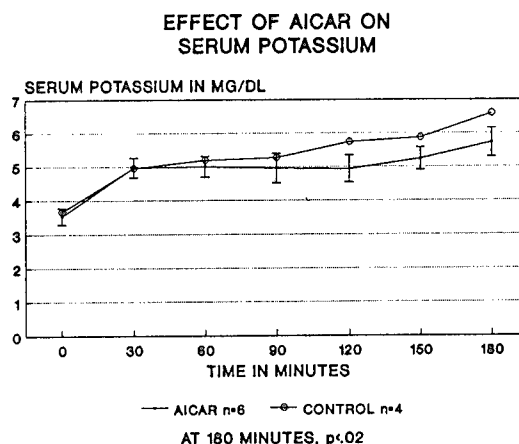EFFECT OF AICAR ON
SERUM PHOSPHATE

SERUM PHOSPHATE IN MG/DL

TIME IN MINUTES

— AICAR n=6   —⊙— CONTROL n=4
AT 200 MINUTES, p<.02
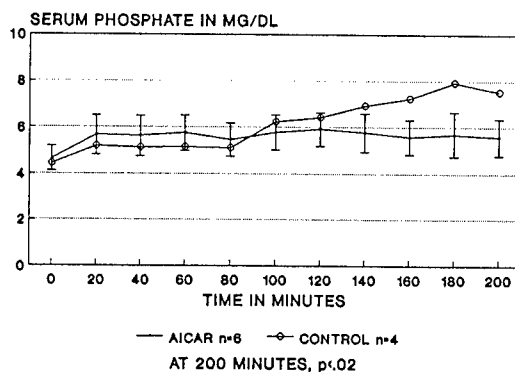
EVANS, WIESMANN

Effects of hemorrhage on potassium are shown in Figure 10. AICA-riboside treated animals had an increase of 2.0 ± 0.38 mEq/l as compared to 3.1 ± 0.42 mEq/l in AICA treated rats (p<.02).

Figure 10

EFFECT OF AICAR ON
SERUM POTASSIUM

SERUM POTASSIUM IN MG/DL
TIME IN MINUTES

— AICAR n=6   —⊕— CONTROL n=4
AT 180 MINUTES, p<.02

At 90 minutes, the AICA-riboside group was significantly less acidotic (pH = 7.378 ± 0.01) than the AICA group (pH = 7.23 ± .03, p<.05). This was not significant at other time intervals.

DISCUSSION:

We have used a severe model of hemorrhagic shock to test the ability of the drug AICA-riboside to prolong survival and improve certain metabolic parameters. This model corresponds to a 40-50% blood loss, an immediate life threatening condition in the human. Metabolic and hemodynamic decompensation are closely interrelated at this degree of hemorrhagic shock. Since low flow states caused by hypovolemia cause ischemia and anoxia with resultant loss of organ function[25], our hypothesis that AICA-riboside has a protective effect and prolongs the survival period should also be reflected in improved metabolic parameters such as higher glucose and improved utilization of phosphate. Our data suggests that an improved metabolic capacity does take place, and is largely occurring in the liver.

Experiments in our laboratory suggest that rising serum inorganic phosphorous levels reflect the breakdown of intracellular phosphorylated nucleotides and the failure of intracellular phosphorylation. Bone, a major storage tissue for phosphorous was an unlikely source of the increased serum levels we measured during shock. Fluctuations in ionized

EVANS, WIESMANN

calcium (data not shown) which could stimulate parathyroid hormone to release calcium and phosphorous from bone were not observed. To evaluate the possibility that renal mechanisms played a role in phosphate kinetics, we performed bilateral nephrectomies in 4 animals and followed serum phosphorous over the same time course as the shock model (data not shown). No changes in phosphorous levels occurred suggesting that an acute loss of renal function did not contribute to the rise in phosphorous.

In shocked nonresuscitated rats, phosphorous levels rose quickly during the compensatory phase reaching 66% of their total increase by "Max Out". Even when shocked for only 6 minutes and then resuscitated, phosphate levels increased 57% over 5 hours. Other investigators have demonstrated that elevated inorganic phosphorous levels in unshocked rats will decrease the degradation of ATP[26]. In the shocked state, however, neither adequate levels of adenylate nucleotide precursors nor cellular energy are available to synthesize ATP. Serum phosphorous, therefore, is a metabolic parameter indicative of metabolic decompensation which occurs earlier than hemodynamic decompensation.

AICA-riboside treated animals demonstrate 2.6 fold longer survival compared to controls. This is manifested by the ability to maintain blood pressure longer during the decompensatory phase of the model. A possible mechanism for improved survival may be reflected in lower serum phosphorous levels in the AICA-riboside group beginning at 90 minutes. Serum phosphate may be taken up more rapidly by cells and incorporated into adenine nucleotides. AICA-riboside which is membrane permeable may provide a substrate for intracellular phosphorylation and conversion to IMP. This mechanism has been previously demonstrated in ischemic cardiac and skeletal muscle[19]. In addition there may be slower breakdown of ATP secondary to improved metabolic function. More ATP may be available to support gluconeogenesis longer during shock. This concept is supported by the higher glucose levels in the AICA-riboside treated animals over a four hour period. Lactate levels (data not shown), however, do not decrease as glucose levels increase. This may be a reflection of the large size of the lactate pool, which is not significantly reduced by the modest increase in gluconeogenesis.

Both D-ribose and L-ribose treated animals demonstrate significantly worse survival times compared to the AICA-riboside group. The longer decompensatory time in the D-ribose group compared to its isomer is not statistically

EVANS, WIESMANN

significant. Some differential benefit might be expected since the L-ribose form of the stereoisomer is not incorporated into purine nucleotides. D-ribose can be phosphorylated to PRPP, and utilized in the _de_ _novo_ and salvage biosynthetic pathways[22].

Administration of AICA-riboside beginning when blood pressure reached 40 mmHg as opposed to waiting until "Max Out", lengthens the compensatory phase 1.7 fold. This longer time is supported by an increased percentage of predicted maximum output that was achieved. This slight hemodynamic benefit may be a reflection of improved vascular tone, although AICA-riboside did not cause changes in blood pressure and it has no effect on cardiac output[19].

Lessened insult at the cellular level is suggested by improved pH and lower serum potassium levels. Arterial blood gas analysis supports an acidosis of metabolic origin, with cell damage leading to intracellular accumulation of hydrogen ions. Potassium which leaks out of cells to balance the charge fluctuations reflects the degree of acidosis. This metabolic data further supports improved cellular integrity during hemorrhage in the AICA-riboside group.

AICA-riboside is an efficient precursor to nucleotide resuscitation and is unique in its biochemical behavior[19-22]. From an energy standpoint, AICA-riboside stimulates the synthesis of a new IMP molecule that requires the investment of only one phosphate bond. This is one fourth the energy used in normal _de_ _novo_ synthesis. AICA-riboside provides an additional advantage over purine salvage mechanisms by not using 5-phosphoribosylpyrophosphate in its pathway, a scarce substrate during shock[19-22]. This metabolic approach is unique in hemorrhagic shock research.

Other advantages over previously studied interventional methods include AICA-ribosides' lack of detrimental effects on the cardiovascular system. The low dose AICA-riboside used in our studies does not cause changes in heart rate, systemic blood pressure, cardiac output or regional myocardial blood flow[19]. As a result, the metabolic benefits we demonstrated during hemorrhage are not compromised by alterations in hemodynamic response. This novel approach to nucleotide resuscitation may prove to have clinical efficacy in improving survival in soldiers with hemorrhagic shock.

EVANS, WIESMANN

REFERENCES

1. Bellamy, R. Wound data and munitions effectiveness team. In Textbook of Military Medicine, edited by D. Jenkin. In Press.

2. Goris, RJ and J. Draaisma. Causes of death after blunt trauma. J. Trauma 22:141, 1982.

3. Douglas, R.G., and J.H.F. Shaw. Metabolic response to sepsis and trauma. Br. J. Surg. 75:115, 1989.

4. Pearce, F.J., R.J. Connett and W.R. Drucker. Phase related changes in tissue energy reserves during hemorrhagic shock. J. Surg. Res. 39:390, 1985.

5. Pearce, F.J., P.R. Weiss, J.R. Miller and W.R. Drucker. Effects of hemorrhage and anoxia on hepatic gluconeogenesis and potassium balance in the rat. J. Trauma, 24:312, 1983.

6. Pass L.J., P.R. Schloerb, F.T. Chow, M. Graham, F.J. Pearce, M.W. Franklin and W.R. Drucker. Liver ATP in hypoxia and hemorrhagic shock. J. Trauma. 22:730, 1982.

7. Melo, L., L.V. Bacalyo and L.D. Miller. Defective oxidative metabolism of rat liver mitochondria in hemorrhagic and endotoxin shock. Am. J. Physiol. 220:571,

8. Hems, D.A. and J.T. Brosnan. Effects of ischemia on content of metabolites in rat liver and kidney in vivo. Biochem J. 120:105, 1970.

9. Baue, A.E., M.M. Sayeed, I.H. Chaudry and M.S. Wurth. Cellular alterations with shock and ischemia. Angiology. 25:31, 1974.

10. Friedman, S.G., F.J. Pearce and W.R. Drucker. The role of blood glucose in defense of plasma volume during hemorrhage. J. Trauma. 22:86, 1982.

11. Pearce, F.J. and W.R. Drucker. Glucose infusion arrests the decompensatory phase of hemorrhagic shock. J. Trauma. 27:1213, 1987.

12. Stoner, H.B., D.F. Heath, O.H. Collins. The metabolism of ^{14}Cglucose, ^{14}Cfructose and 2-^{14}Cpyruvate after limb ischemia in the rat. Biochem. J. 76:135, 1960.

EVANS, WIESMANN

13. Chick, W.L., R. Weiner, J.Cascarano and B.W. Zweifach. Influence of Krebs-cycle intermediates on survival in hemorrhagic shock. Am. J. Physiol 215:1107, 1968.

14. Ida T, M. Ukikusa, M.Yamamoto, A. Nakase, K. Ozawa and R. Tobe. Stimulatory effects of adenosine on hepatic adenine nucleotide and energy charge levels in shocked rats. Circ Shock 5:383, 1978.

15. Talaat, S.M., W.H. Maisson and J.A. Schelling. Effects of adenosine triphosphate administration in irreversible hemorrhagic shock. Surgery 55:813, 1964.

16. Chaudry, I.H, M.M. Sayeed and A.E. Baue. Effect of adenosine triphosphate-magnesium chloride administration in shock. Surgery 75:220, 1974.

17. Hirasawa, H., S. Oda, H. Hayashi, Y. Ohtake, M. Odaka, and H. Sato. Improved survival and reticuloendothelial function with intravenous ATP-MgCl$_2$ following hemorrhagic shock. Circ Shock 11:141, 1983.

18. Chiao, J.J., J.P. Roberts, H.P. Illner and G.T. Shires. Permeability of red cell membranes to adenosine triphosphate (ATP) molecules during hemorrhagic shock. Surgery 102:528, 1987.

19. Sabina, R.L., K.H. Kernstine, R.L. Boyd, E.W. Holmes and J.L. Swain. Metabolism of 5-amino-4-imidazole carbo- xamide riboside in cardiac and skeletal muscle. J. Biol Chem. 257:10178, 1982.

20. Flanagan, W.F., E.W. Holmes, R.L. Sabina and J.L. Swain. Importance of purine nucleotide cycle to energy pro- ction in skeletal muscle. Am. J. Physiol 251:C795, 1986.

21. Swain, J.L. J.J. Hines, R.L. Sabina O.L. Harbury, and E.W. Holmes. Disruption of the purine nucleotide cycle by inhibition of adenylosuccinate lyase produces skeletal muscle dysfunction. J. Clin. Invest. 74:1422, 1984.

22. Mauser, M., H.M. Hoffmeister, C. Nienaber and W. Schaper. Influence of ribose, adenosine and "AICAR" on the rate of myocardial adenosine triphosphate synthesis during reperfusion after coronary artery occlusion in the dog. Circ Res 56:220, 1985.

23. Wiggers, C.J. Physiology of shock. New York: Oxford University Press, 1950, p 121-146.

EVANS, WIESMANN

24.   Connett, R.J., F.J. Pearce and W.R. Drucker.
       Scaling of physiological responses: a new
       approach for hemorrhagic shock. Am. J. Physiol.
       250:R951, 1986.
25.   Pearce, F.J., P.R. Weiss, J.R. Miller and W.R.
       Drucker. Effect of hemorrhage, anoxia and graded
       ischemia on hepatic gluconeogenesis. Surg Forum.
       p 1-5, 1982.
26.   Johnson, M.A., K. Tekkenat, S.P. Schmalty and
       I.H. Fox. Adenosine triphosphate turnover in
       humans. J. Clin Invest. 84:990, 1989.

AUTOMATED ANALYSIS OF INTERACTIONS BETWEEN
TERRAIN AND GROUND FORCE MOBILITY

TERRIL C. FALLS, Mr.
*ROBERT B. UNDERWOOD, III, CPT
U. S. Army Engineer Waterways Experiment Station
Vicksburg, MS 39180

INTRODUCTION

The Condensed Army Mobility Model System (CAMMS) enables commanders to rapidly analyze the interactions between terrain and ground force mobility. The need to continually assess the effects of terrain on ground force mobility is fundamental to a commander's decision making. By automating the analysis of these interactions, the commander's decision cycle can be greatly compressed, and his orders can be based on more data than would otherwise be possible. Providing the needed analytical capability requires the integration of current hardware capabilities with advanced prediction models, digital terrain data bases, and user-friendly graphical interfaces. This integration has been accomplished over time in conjunction with an active demonstration program: CAMMS has been constantly been demonstrated to Army units world-wide. These demonstrations have provided the critical feedback necessary to meet the Army's needs and to gain acceptance. This paper describes how CAMMS augments the terrain-mobility evaluation process.

History of CAMMS

CAMMS has evolved from the Army Mobility Model (AMM), which was codeveloped by the U. S. Army Engineer Waterways Experiment Station (WES) and Tank-Automotive Command to analyze the mobility performance of conceptual combat vehicle designs. The AMM incorporated fundamental vehicle-terrain interaction relations that were formulated from vehicle field testing programs and laboratory work begun in the World War II era. The NATO Reference Mobility Model (NRMM) uses AMM as the cornerstone of its mobility prediction capability for the NATO materiel development community. The same vehicle mobility prediction algorithms are used in AMM, NRMM, and CAMMS, giving the Army a robust and consistent mobility prediction capability for both materiel development and for tactical usage.

The initial CAMMS was developed by WES in 1983, providing a simplified version of the AMM suitable for a microcomputer (Apple II). The initial CAMMS could make cross-country mobility predictions for military vehicles and display those predictions graphically. As microcomputer capabilities advanced, it became feasible to make CAMMS more powerful. CAMMS was modified to run on the MS-DOS and VMS operating systems. As the user community began to appreciate CAMMS' potential, more capabilities were developed. New algorithms were incorporated for on-road speed prediction, tactical bridging, obstacle effects, and weapons coverage. Output was enhanced for better screen displays and higher-resolution printing. Three-dimensional views were developed; now the user can get a "bird's eye" view or a view of the terrain from a battle position.

CAMMS Software and Hardware

CAMMS software consists of the various programs that compute the Tactical Decision Aids (TDA). The actual mobility models account for about thirty percent of the total program system code. Software written to handle graphics, database management, and the user interface menus accounts for the remainder of the system code. The majority of the program coding has been written in the C language, with the bulk of the remainder written in FORTRAN. Some routines requiring fast execution have been written in assembly language to reduce user wait times.

CAMMS runs on MS-DOS and VMS-based computer systems. The following hardware is the minimum required to run the MS-DOS based CAMMS software:

286 or 386 central processor unit
MS-DOS 3.2 or later
Math coprocessor
EGA or VGA graphics monitor
20MB free hard disk space
Epson-compatible black-and-white or color dot matrix printer

A "mouse" pointing device may be optionally used to operate the software.

CAMMS Capabilities

CAMMS is fundamentally a collection of mobility-related TDA's. These TDA's have been incorporated into the system at the suggestion of the user community: the Army in the field. The major CAMMS TDA's are:

Cross-country speed

On-road speed
Foot soldier speed
Maneuver damage prediction
Unassisted stream crossing potential
Fixed bridge crossing potential
Tactical bridge emplacement
Interactive route evaluation
Obstacle emplacement and effectiveness
Weapons lines of sight
3-d perspective views
Map image projection

These TDA's can be displayed on a graphics monitor and can be plotted on a variety of printing devices.

Mobility Prediction Considerations

Data

Mobility is directly influenced by vehicle characteristics, terrain factors, weather conditions, and natural and man-made obstacles. CAMMS takes these factors into consideration when predicting vehicle speed. The pertinent mobility factors are stored as data in several CAMMS databases.

The data requirements for the terrain database are the most extensive of the various CAMMS databases. The digital terrain descriptions are divided into three major categories: cross-country, roads, and drainage. The cross-country data are stored as raster data typically with one hundred meter resolution. The cross-country data include the following terrain factors: elevation, slope, vegetation, land use, soil type, and natural obstacles. The road descriptions are stored as vector data and includes elevation, slope, surface material, curvature, superelevation, and road side vegetation. The drainage data are also stored as vector data and includes stream width, stream depth, water depth, bank angles, and soil type of the stream bottom. The current shortage of available digital terrain data limits the application of CAMMS, but this limitation should be lessened with the planned production of Interim Terrain Data (ITD) by the Defense Mapping Agency.

The vehicle database includes the vehicle geometry, inertial characteristics, powertrain, and wheel or track characteristics. An extensive library has been accumulated of vehicle characteristics of U.S. and foreign military vehicles.

Weather data can be used by the system in two ways. A historical database of statistical precipitation levels is maintained by the system for specific geographic regions. Or, the user can monitor rain gauges within his area of interest and the system will account for the accretion and depletion of soil

moisture through CAMMS' Soil Moisture Strength Prediction (SMSP) model. The SMSP Predicts soil strength according to current or statistical weather.

CAMMS Submodels

Soil Moisture Strength Prediction (SMSP) Model

The SMSP model provides the means of using actual or historical precipitation for mobility predictions. SMSP uses the precipitation, soil type , and drainage characteristics of the terrain to calculate the moisture content of the soil. Based on the moisture content and soil type, the soil's strength (that is, its ability to sustain traffic) may be calculated. This is fundamental to the making of a mobility prediction. In order to maintain accurate soil moisture strength data, the SMSP model must be updated and run each day. However, monthly averages may be used if rainfall data are not available to the user.

Mobility Models

The CAMMS cross-country and road mobility prediction models are designed to predict the unassisted steady-state mobility performance of a vehicle. The models calculate the forces acting on the vehicle due to the terrain and the force generated by the interaction of the powertrain of the vehicle and the surface upon which the vehicle operates. A comparison of these forces is used to calculate the maximum speed of the vehicle on the terrain. Various other speeds are also calculated which can limit the speed of the vehicle. Visibility limited speed is calculated using driver visibility and braking characteristics of the vehicle with the premise that the driver will limit the speed of the vehicle, such that he maintains the ability to recognize an obstacle and stop the vehicle. A micro-roughness description of the terrain and the vehicle's ride characteristics are used to calculate a speed, which limits the shock and vibrations absorbed by the driver to within acceptable limits. Vegetation and obstacle spacing are used to calculate the speed at which the vehicle can maneuver around the vegetation or obstacle, if unable to override the vegetation or obstacle. Road curvature is used to calculate the maximum speed for operation on the road. A comparison of these various speeds yields the speed prediction for the vehicle operating on the described terrain. This process is used to calculate the speed of the vehicle for up-slope, down-slope, and level operations. The up-slope, down-slope, and level speeds are used to calculate and average speed assuming the vehicle operates for an equal distance in each direction. Outputs from the model are the average speed, the three directional speeds, and the reason code which describes what is limiting the vehicle's speed. The average speed is used for display purposes and the route evaluation uses the directional speeds (up, down, and level) to calculate route performance. The reason code is used to assess requirements for engineering support to traverse the described terrain.

The river crossing model is designed to predict the ability of the vehicle to cross a given river unassisted and, if crossing is possible, provides the approximate crossing time and speed. The river crossing prediction model considers the following factors in determining the ability of the vehicle to cross the river: traction available for entering the river, traction required to exit the river, interference of the body of the vehicle with the river banks or bottom, the ability of the vehicle to ford or swim the water in the river and the vehicle's ability to cross vegetation on the river banks. Outputs from the model are GO-NOGO for crossing for crossing the river and the time required for crossing, if possible, or the reason code which describes why the vehicle could not cross, if not. As was the case for the cross-country and road output prediction, the GO-NOGO prediction is used for display purpose, the crossing time is used to calculate route performance, and the reason code is used to assess engineering support requirements for the crossing.

Terrain Factor Analysis

CAMMS provides the capability to display a color-coded map of the basic terrain factors which are stored in the terrain data base. In addition to the basic terrain factor display, CAMMS provides the capability to provide special purpose display products which are a combination of the terrain factors. The special products are produced by searching the data base for terrain descriptions which meet given criteria for the product and generate a color coded map display to portray the results.

APPLICATIONS

CAMMS has been used in numerous field exercises, most recently during REFORGER 90 as part of U. S. Army, Europe's Digital Terrain Enhancement Program (DTEP). The DTEP was initiated by U. S. Army, Europe to provide terrain visualization capabilities to Brigade commanders beginning with REFORGER 90. DTEP consists of hardware, software, and database integration.

REFORGER 90 was conducted over a maneuver area extending roughly from Stuttgart to Munich. The US VII and V Corps were antagonists, with VII Corps on the west and V Corps on the east. Both Corps had DTEP systems down to Brigade level, although not all Brigades had systems. Technical support and assessment was provided to both sides.

. DTEP hardware had several different configurations, primarily personal computer-based, running the MS-DOS operating system. Desktop computers were used at Corps and Division levels, with two at Brigade level. Laptop computers were used primarily at Brigade level. The Digital Topographic Support System Prototype (DTSS-P), which is a Microvax computer running on the VMS operating system, provided DTEP capabilities at the Theater and Corps

levels. The U. S. Army Forces Command's Automated Intelligence Support System (FAISS), an MS-DOS system, was fielded at Division and above. A variety of output devices was fielded, including color display monitors, color and black-and-white dot matrix printers, electrostatic color plotters, and pen plotters for the DTSS-P.

DTEP analysis capability was provided by three software packages: the CAMMS, Terrabase written by the U. S. Military Academy, and the commercial program Desktop Mapping System (DMS). All three packages provided unique capabilities, although inevitable overlapping of capabilities existed. Not all software ran on all hardware; operating system compatibility, output devices and disk storage space constrained universal installation. CAMMS was unique among the software in being installed on all hardware configurations.

DTEP digital terrain databases were created for the REFORGER maneuver area to support the software packages. This necessitated the creation of three different databases for CAMMS, Terrabase, and DMS. The 649th Engineer Battalion (Topographic) used Terrabase to scan Tactical Terrain Analysis Data Bases (TTADB) for Terrabase and CAMMS, and digitized road and drainage data for CAMMS. WES processed digitized factor data into CAMMS databases. German-provided Standardized Raster Graphics data were also incorporated into CAMMS databases. Defense Mapping Agency produced Digital Topographic Elevation Data (DTED) was used by all software packages.

The primary terrain visualization and analysis capabilities used by the units were CAMMS-generated. Terrabase was used extensively for line-of-sight analyses. DMS was essentially not used. Most used were:
1. Mobility overlays (CAMMS)
   a. cross-country
   b. road-movement
   c. obstacles
   d. unassisted river crossing
2. River crossings (CAMMS)
   a. fixed bridges
   b. tactical bridges
   c. ford sites
3. Lines-of-sight (Terrabase)
   a. antennas
   b. ground surveillance radar
   c. weapons sighting
4. 3-Dimensional views (CAMMS and Terrabase)
   a. bird's eye views
   b. view from battle position
   c. view along a flight path
   d. rotational views (Terrabase)
5. User-defined TDA's (CAMMS with technical assistance)
   a. tactical assembly areas

538

b. potential artillery positions

The DTEP exercise showed that additional capabilities are needed, including communications capabilities, improved database management, more powerful mobility algorithms, and improved user interfaces. The system must have ability to transmit updated terrain factor files higher, lower, and to adjacent units. The user must be able to access any subset of the terrain database at will. This was not possible with the CAMMS data produced for REFORGER 90, and this resulted in suboptimal CAMMS use.

The DTEP system must incorporate unit movement and avenues of approach generation capabilities to fully support Intelligence Preparation of the Battlefield conducted by S-2's and G-2's. This frequently requested capability is being developed under an active Research, Development, Test, and Evaluation work unit. The DTEP system must enable users to switch between grid reference systems: the Military Grid Reference System is most commonly used by combat units.

Commanders and staffs want the capability to produce operational overlays before anything else. Manual preparation of operational overlays expends an inordinate amount of precious time. Direct terrain data query and data update will enable users to directly access the terrain database by pointing to a terrain feature and having the system describe that feature. Users also want to be able to add, delete, or modify feature descriptions as reconnaissance reports come in. This requirement of course necessitates a communications capability for the system.

Commanders and their engineers consider man made obstacles simultaneously with other mobility terrain factors. This necessitates an obstacle plan management capability that provides the ability to display obstacle locations and status and to communicate obstacle locations, types and status to other units. The capability to accurately predict obstacle effects on unit maneuver is also needed.

Considerations for the Future

The acceptance of mobility-terrain interaction software as an Army terrain visualization and analysis tool requires the resolution of several issues. These issues, which have broad implications, center around data availability, hardware and software capabilities, training, and maintainability.

Digital terrain data availability and accuracy are key to the usefulness of any terrain visualization and analysis system. Users must have a high degree of confidence in the fidelity of the digital map which presents a great challenge to the producers of tactical terrain data. Data coverage must extend beyond areas under friendly control into unfriendly territory, this compounding the data acquisition and validation problem. Detailed road,

stream, and bridge data are demanded by the combat units to support their battlefield analyses.

Users also demand the ability to update and maintain their own terrain databases in order to reflect the current state of the battlefield. This requirement necessitates a sophisticated database management and communication system. Current methods are manual, and given the enormous amount of data generated by the conduct of warfare, these manual methods are totally inadequate. The information effectively does not get transmitted, adding greatly to the fog of war.

Further hardware development is needed to adequately support mobility analysis software. High speed, large display graphics capabilities are essential to allow products to be generated within the commander's decision cycle. At the same time, hardware must be rugged, compact, durable and maintainable, just like all other Army fielded equipment. Uninterruptable power supplies must be robust and tolerant of dirty power sources.

CONCLUSIONS

1. WES and the terrain analysis community have gained valuable first-hand insights into the needs of the Army in the field for automated terrain-mobility analysis capabilities. This experience will pay big dividends in producing more effective software.

2. Software capabilities must support staff activities. Battle staffs will not dramatically change their operating procedures to use a software system. Therefore, the software must be designed to support and augment what various staff sections already do and the order and methods in which they perform their tasks.

3. The value of the information generated by automated terrain analysis systems is severely limited if there is no automated method of transmission. The transmission capability of the Army's tactical communications systems can constrain the ultimate usefulness of automated terrain analysis products.

4. Software must be self-trainable and self-explanatory. Training must be minimized if the software is to be used. Therefore, the software must be so easy to use that a soldier can learn it at the keyboard without attending a class or reading a user's manual.

5. Software must be able to tell the user how an answer was derived. The first question from a user upon seeing a product such as a mobility map is "How is this so?". They will not believe the answer until it is explained to them. The software must be able to explain any answer at any time.

Acknowledgements

Breakdown in Silicon p-n Junctions due to Short Pulse, High-Field Electromagnetic Stress (U)

Christian Fazi,* Dr. and Alford Ward, Dr.
U.S. Army Harry Diamond Laboratories
2800 Powder Mill Road
Adelphi, Maryland 20783-1197

Roger C. Westgate, Dr.
Electrical Engineering Department
Johns Hopkins University
Baltimore, MD 21218

Introduction

While working for Harry Diamond Laboratories some years ago, we were asked to determine the electromagnetic compatibility of a communication system. Since the laboratory had just obtained some of the fastest instrumentation for generating and recording rf waveforms, it was decided to extend all measurements of transient response down to 10 ns. Before this effort, most of the interest in transient response was limited to 0.1 µs. We were surprised when certain silicon junctions appeared to fail so rapidly that measuring their response to the shorter transients was not possible. Further, the responses did not scale according to any of the known thermal failure laws. (This is just as startling as someone making far-field, free-space power-density measurements and observing other than the $1/R^2$ falloff). Even more surprised was a supervisor who, after allocating more than a man-year of effort to generate a comprehensive software package to calculate all the pertinent system response parameters from measured data, was told that the communication gear failed with practically no energy. While the new measurement setup was blamed, we felt that something unusual might be occurring and that the existing failure models were invalid on these time scales. At the time there was no mention of this problem in the literature.

Recently there has been considerable interest in very short pulse applications, resulting in more studies in source development and device effects. The fields most impacted are impulse radar, frozen wave generators, and transient electromagnetic effects. Many new systems are required to operate at higher frequencies and at higher peak powers, stressing semiconductor devices to their limits. This has resulted in an increased emphasis on semiconductor failure studies, especially in the area between first and second breakdown of bipolar junctions.

The mechanisms for initiating second breakdown failure in an epitaxial p-n junction, which occur at higher bias levels beyond first breakdown, have been reported by many authors[1-12] and are understood for simple geometries with areas larger than 10 µm^2 and for pulses longer than 0.1 µs. The theory for studying semiconductor failure is based on several high-field, nonlinear models that have been developed over the years and are supported by experimental data. These types of models fall into two categories, thermal[4,6] and electrothermal.[1,13] However, when high-field short pulses lasting less than 50 ns are injected into a p-n junction, the failure mechanism often does not appear to exhibit any thermal dependence, and it further appears that the junction fails without ever exhibiting second breakdown. This type of failure has been named "current mode breakdown" or "instantaneous breakdown." Second breakdown, which occurs after avalanche breakdown (or first breakdown), is an energy-dependent type of failure, while current mode breakdown appears to have properties similar to dielectric breakdown, an electric-field amplitude-dependent failure.

The objective of this paper is to determine which device parameters lead to this nonconventional failure in silicon bipolar junctions. To accomplish this, experimental data and a one-dimensional drift-diffusion model are used to obtain the current density versus applied voltage and to determine the point of negative differential resistance for different types of p-n junctions. The span of device parameter variations is as wide as the many different silicon technologies dictate: 10^{14} to 10^{17} cm^{-3} for doping densities and epitaxial thicknesses from 3 to 15 µm. This study also contributes new experimental data by extending the existing failure data below 100 ns; data at this level, which presently are very scarce, will help us understand the response of p-n junctions to high-bias excursions of short duration.

Second Breakdown in Semiconductor Devices

Second breakdown, a precursor to burnout in p-n junctions, is often used as a means to determine the maximum safe operating level of bipolar semiconductor devices. Many devices operate close to this limit, due to either fast switching speed and high power handling requirements or to exposure in a transient electromagnetic environment. The concept of a pulse-width-dependent minimum damage energy (or power) of a p-n semiconductor defines a safe operating region that can influence pulsed device design. This can be accomplished when there is a quantitative understanding of the relationship between device parameters and failure. Although many bipolar semiconductor manufacturers include this information in their device catalogs, they usually do not test with pulses less than 10^{-3} to 10^{-4} s. Because of the large parameter variations among the different p-n junctions, it is not feasible to obtain the safe operating limits for every device type by statistical testing alone. This type of testing cannot improve the understanding of the failure mechanisms without some device physics modeling capability.

Second breakdown was observed 30 years ago by Tauc and Abraham[2] in bipolar semiconductors. This behavior occurs in a reverse biased p-n junction when the bias excursion is taken beyond avalanche or Zener breakdown. The junction current versus voltage dependence for the two bias modes is shown in figure 1. It is experimentally observed as a sudden current jump, accompanied by a voltage drop, as shown in figure 2. Chen et al[3] report that second breakdown takes place in three stages: *nucleation*, *growth*, and *melt*. Nucleation exists when the local temperature in the junction reaches some critical value. When this temperature exceeds the peak resistivity-temperature of the lightly doped side of the junction, the current is no longer limited by the positive temperature coefficient of resistance, and thermal runaway occurs. This dependence for silicon is shown in the resistivity versus temperature curve of figure 3. A certain delay time is observed before nucleation and growth of the current filament. This is due to the time required to heat the silicon material. The largest voltage drop seen in typical second breakdown oscillograms shows the effect of the melt stage, as illustrated in figure 4. This enhanced voltage waveform is taken with a constant current source.

The Thermal Model

Since the electrical characteristics of p-n junctions are strong functions of temperature, the simplest model is based on the heat flow equation. The thermal dependencies are obtained by solving the heat flow equation (1.1), which is a parabolic partial differential equation, subject to the Neumann boundary conditions, shown in figure 5:

$$\rho c \frac{\partial T}{\partial t} = |\mathbf{J} \cdot \mathbf{E}| + K_T \frac{\partial^2 T}{\partial^2 x} + \frac{\partial K_T}{\partial T}\left(\frac{\partial T}{\partial x}\right)^2 , \qquad (1.1)$$
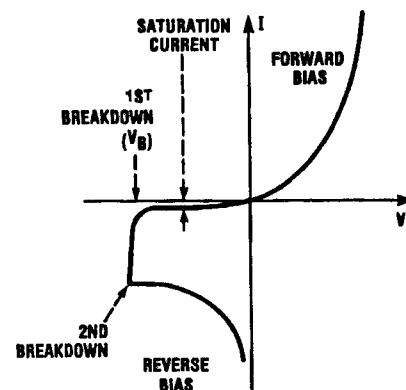

Figure 1. Forward and reverse current versus voltage dependence for a p-n junction.
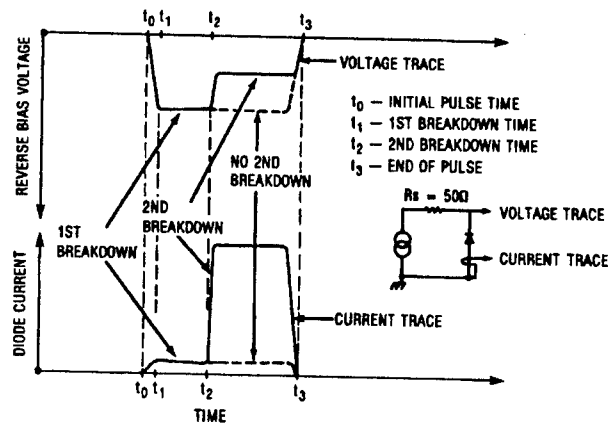
544

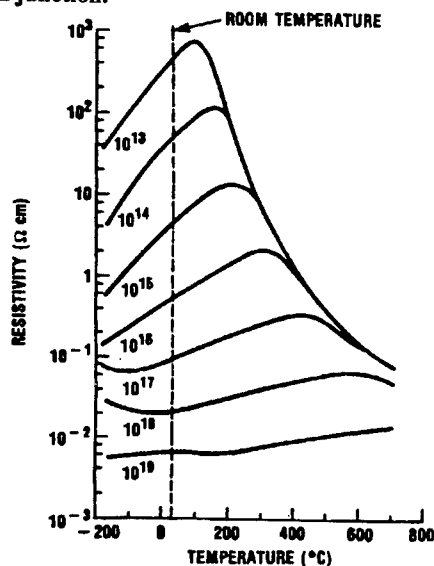Figure 2. Voltage and current second breakdown traces for a p-n junction.

Figure 3. Variation of resistivity with temperature for doped silicon.
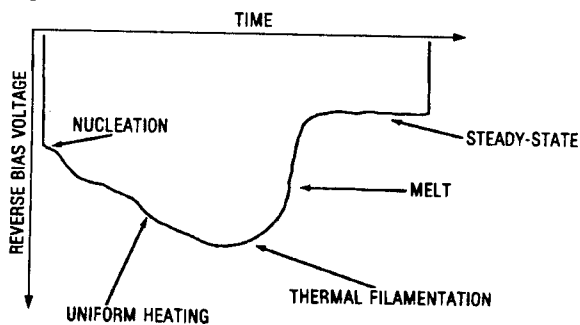

Figure 4. Thermal second breakdown voltage in a p-n junction under constant current regime.

where ρ is density, c is the specific heat, and K_T is the thermal conductivity of silicon. The term $Q = |\mathbf{J} \cdot \mathbf{E}|$ represents the heat generation rate, where $\mathbf{E}$ is the electric field vector of the junction and $\mathbf{J}$ is the current density vector flowing through it.

Of the basic models available used to predict the minimum damage energy of bipolar semiconductors, the simplest is a one-dimensional thermal model[6] based on a semi-infinite semiconductor medium, with a line heat source along the junction boundary. This approach cannot model electrical instabilities and ignores space charge effects in the junction. Also it cannot distinguish an abrupt junction from a linearly graded one. The model assumes that the current is injected uniformly over the active area and that the heat generated in the junction region increases the temperature of the device until second breakdown occurs. This assumption sets the power to fail at a critical temperature which can be any temperature depending on the modeling requirements, such as the melt, eutectic, or intrinsic temperature of the semiconductor material, and relates the power to fail with the failure delay time. Only simple device parameters such as thickness, thermal conductivity, and heat sink temperature are needed. The thermal model is reviewed here for the simplest case where the active area and chip area of the p-n junction are equated. For the one-dimensional case the average temperature increase in the junction for a given power level is given by Minniti:[6]

$$\Delta T = \frac{2P}{\alpha l A K} \sum_{n=1}^{\infty} \left(1 - e^{-\beta^2 k t}\right) \frac{\sin \beta l \sin \beta \alpha \sin \beta x}{\beta^3}, \quad (1.2)$$

where

α = the heat source thickness (also the depletion thickness),

P = the power absorbed in watts,

β = $\dfrac{(2n-1)p}{2l}$ $(n = 1 ... \infty)$, $\quad$ (1.3)

k = the thermal diffusivity of silicon (29.5 μm^2/■ s),

K = the thermal conductivity of silicon (5.2 10^{-5} W/μm K),

l = total thickness of the semiconductor (excluding heat sink),

A = the chip area in μm^2, and

ΔT = change in temperature.

545

This solution has a temporal and spatial dependence, relating the heat generation rate to the temperature buildup in the slab. Assuming that an initial failure occurs at the eutectic temperature for Si-Al, which is below the silicon melt temperature, 1685 K, it is possible to solve for the power to fail by fixing the value of $\Delta T = T_{eut} - T_{sink} \sim 350$ K. The asymptotic solution for equation (1.2) is then given in figure 6. The thermal model does identify three time-dependent thermal regions: the thermal adiabatic region, the transition region, and the thermal nonadiabatic region. In the adiabatic region, power to fail shows a t^{-1} dependence, where "t" is the pulse width time. In the transition region, power to fail shows a $t^{-1/2}$ dependence, while in the nonadiabatic region it is constant or has a t^0 dependence. Integrating power to fail over time in the same regions of interest yields the energy to fail time dependence. When current mode type failures occur, they appear only in the adiabatic region as shown in figure 6, and cannot be predicted by this model.

Current mode failures are more commonly observed in three- and four-layer bipolar structures such as transistors and silicon controlled rectifiers (SCRs), rather than in diodes, even with pulses with slower risetimes. This type of failure is attributed to the role the base region has on gain and on space charge effects.[14] With the exception of tunnel diodes and varactors, diodes on the other hand do not exhibit gain and the leading causes of device failure can be different. In order to understand why a current mode type of failure is observed in some types of diodes and not in others, it is necessary to use an electrothermal model based on drift-diffusion theory.

The Electrothermal Model

The electrothermal model[1] is considerably more complex than the one just described. The great disparity between electronic and thermal time scales poses a serious problem in simulating p-n junction burnout. The electronic time scale is limited by the dielectric relaxation time given by the ratio of the dielectric constant and the conductivity of silicon ($\tau = \varepsilon/\sigma$), and for silicon p-n junctions it is less than a picosecond. Thermal time scales are on the order of a microsecond. For an electronic cycle occurring in the nanosecond regime, the temperature may only change slightly, since thermal diffusion is unimportant in this time scale.

From the electronic state equations and Maxwell's equations it is possible to derive the five basic coupled differential equations describing semiconductor behavior. Since this involves macroscopic modeling with no quantum effects, this modeling can be limited to the drift-diffusion, without invoking Boltzmann transport theory. The following assumptions are made: (1) the effects of magnetic fields, due to the diode currents, can be ignored; (2) transient transport phenomena are insignificant (i.e., electronic equilibrium has been reached); and (3) heavy doping phenomena in the junction are excluded (i.e., the
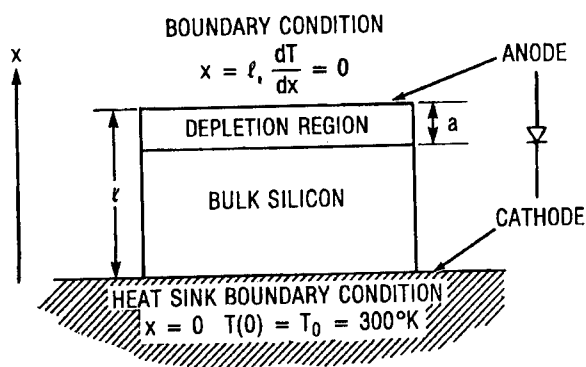

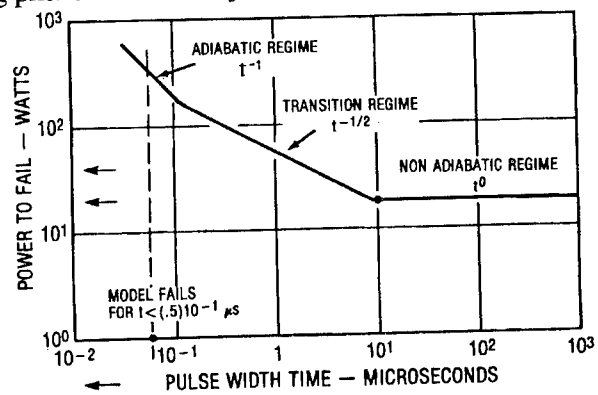Figure 5. One-dimensional thermal model p-n junction.

Figure 6. p-n junction power to fail versus pulse width dependence.

Fermi energy level remains in the gap). Furthermore since the silicon crystal structure is cubic, we also assume that (4) there is sufficiently high symmetry involved that mobility, diffusivity, and permittivity can be regarded as scalar terms instead of tensors.

The electrothermal model then consists of the following stiff, nonlinear partial differential equations, the heat equation (1.1), Poisson's equation (1.4), the electron and hole continuity equations (1.5 a,b), the current density equations (1.6 a,b), and the total current equation (1.7):

Poisson's equation

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{q}{\varepsilon}(n - p + P - N) , \qquad (1.4)$$

Electron and hole continuity equations

$$q\frac{\partial n}{\partial t} = -\left|\frac{\partial J_n}{\partial x}\right| + |\alpha J_n| + |\beta J_p| - R , \qquad (1.5a)$$

$$q\frac{\partial p}{\partial t} = \left|\frac{\partial J_p}{\partial x}\right| + |\alpha J_n| + |\beta J_p| - R . \qquad (1.5b)$$

Current density equations

$$J_n = nq\mu_n E - qD_n\frac{dn}{dx} , \qquad (1.6a)$$

$$J_p = pq\mu_p E + qD_p\frac{dp}{dx} . \qquad (1.6b)$$

Total current density equation

$$J = J_n + J_p + \varepsilon\frac{dE}{dt} , \qquad (1.7)$$

where

ε	=	dielectric constant,
P, N	=	doping density concentrations,
q	=	charge of the electron,
n, p	=	electron and hole mobile carrier density concentrations,
ϕ	=	potential,
α, β	=	electron and hole ionization coefficients,

and

E	=	electric field vector,
J_n, J_p	=	electron and hole current density vectors,
μ_n, μ_p	=	electron and hole mobilities,
D_n, D_p	=	electron and hole diffusion constants, and
R	=	electron-hole recombination rate.

Even though one might wish to solve these equations in three dimensions for all time in analytical form, such solutions rarely exist. Instead, numerical solutions in finite difference form are commonly used. Although numerous models exist, they are all based on the above equations in one form or another. For this investigation the DIODE model by A. L. Ward[1] was chosen.

Fabrication of Test Devices

Depending on the type of application of the semiconductor device, silicon is cut in three different crystal orientations. The (111) orientation is used for bipolar circuits, the (100) orientation for metal oxide structures, since it favors SiO_2 formation, and the (110) orientation, which has strong anisotropy etching characteristics, for devices that need deep isolation. For the test devices in this study the (111) orientation was chosen, since we are restricting our study to bipolar technologies. Although p-n junctions can be fabricated several ways, presently, the epitaxial growth is the most common silicon process, which justifies this type of technology for the experimental test devices.

These test diodes, or p-n junctions, are formed by first growing an epitaxial (epi) layer on a heavily doped substrate. The substrate is usually 10 mils thick. Since n-type diodes are faster than p-type diodes,

due to higher mobility and (therefore) conductivity, the substrate and epi layers are chosen to be n-type. Depending on the device application, whether for high or low voltage and analog or digital functions, the epi region can vary from 10^{14} to 10^{17} cm^{-3}. The p contact is made by thermal diffusion or by ion implant. Sometimes both methods are needed when very high doping levels are required. (The symbol p$^+$ is used to denote high concentration). When metal contacts are made, high doping concentration is required. If the doping concentration falls below 10^{18} cm^{-3}, a Schottky barrier is formed instead of an ohmic contact, resulting in multiple junctions connected in series, instead of a single one. The cross-sectional view of the epi process for a junction is shown in figure 7.

Although the epi layer can vary from 2 to 100 μm, thicknesses greater than 20 μm are used for high-voltage applications only and are not typical of the integrated circuit process. The range of 3 to 15 μm is therefore suitable for the experimental test devices. The range of variations of epi concentrations and thicknesses can be viewed in the two-dimensional array shown in table 1 in relation to the major silicon bipolar technologies presently produced.

Design of Test Devices

Six lots were selected to represent the spread of the epitaxial variations. (This number had to be limited because of the high cost of silicon processing). Each lot was assigned a processing identification code and this code was etched next to every p-n junction on the wafer. The code consists of four fields (there is no relationship to any commercial trade designations). Figure 8 is an electron micrograph showing the topology of a circular p-n junction. Designing the six lots required producing the photolithographic masks to control the shape and size of the anode contact made on the epi layer. The wafers were purchased each with six different epi layers forme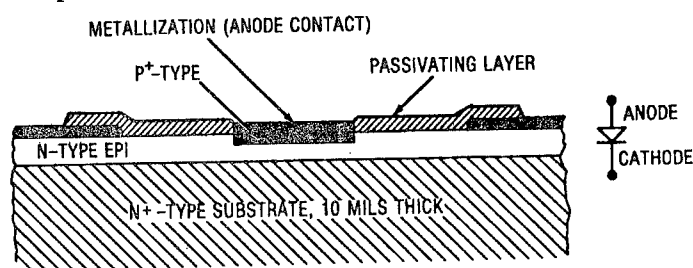d on the substrates. The six completed wafer lots (consisting of 18 wafers in total) and the corresponding processing codes are shown in figure 9. Each wafer consists of 1900 diodes organized in 25 diode arrays. For this research 80-μm-diameter diodes were used, which is a typical size used in commercial devices, such as the very common 1N4148 signal detector diode. The layout needed to produce the photographic masks and reticals was prepared on a GE-Calma system that included standard square pads for contact probing.

To identify the different lots, a special processing code was devised. In the first field of this four-digit code, a "1" or a "0" indicates the silicon crystal orientation in the (111) and (100) directions, respectively. For this case study only the (111) direction is used. In the second field an "N" or "P" indicates the type of epi layer. For this case study only N-epi is used. In the third field a "4," "5," "6," or a "7" indicates a doping concentration of 10^{14}, 10^{15}, 10^{16}, or 10^{17} cm^{-3}. For

Figure 7. Cross-sectional view for an epi p-n junction.

Table 1. Silicon expitaxial parameter array

EPI THICKNESS	EPI CONCENTRATION $\longrightarrow$			
	10^{14}	10^{15}	10^{16}	10^{17}
3 μm	———— ANALOG FUNCTIONS ———— PIN DIODES (1N41)		——— DIGITAL FUNCTIONS ———	FAST DETECTORS (1N71)
7 μm			SIGNAL DETECTORS (1N62)	
12 μm		RECTIFIERS (1N53)		
15 μm	AVALANCHE HV RECTIFIERS (1N44)			ZENER AND OPTICAL DEVICES (1N74)

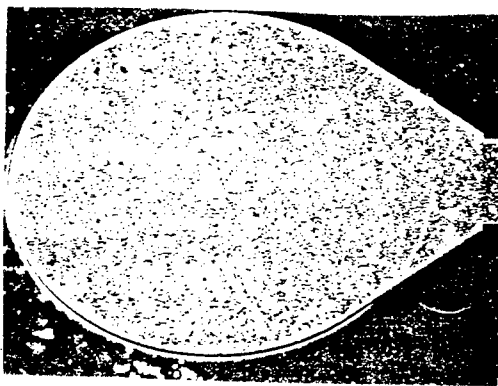548

FAZI, WARD, & WESTGATE

Figure 8. Electron micrograph of a 1N44 circular p-n test junction and contact pad.
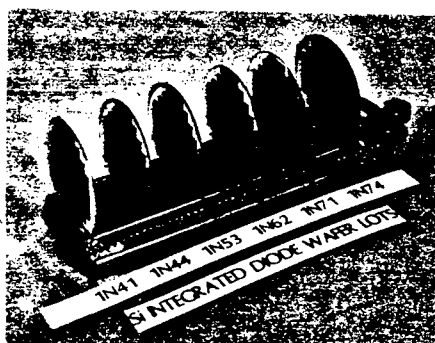

Figure 9. The six test wafer lots.

these test devices, all four doping concentrations are used. The last digit represents an arbitrary epi thickness that varies from 3 to 15 μm ("1" corresponds to 3 μm, "2" corresponds to 7 μm, "3" corresponds to 12 μm, and "4" corresponds to 15 μm). The six designated lots are 1N41, 1N44, 1N53, 1N62, 1N71, and 1N74.

The wafers were profiled with the use of a four probe spreading resistance technique, and the six graphs are shown in figure 10. Knowing these doping profiles is critical for accurate computer modeling results. The most nagging problem in modeling commercial devices has been the need to guess the doping profile of the device, since manufacturers supply only the electrical parameters and nothing more, most often invoking proprietary design restrictions.

p-n Junction Modeling and Failure Criteria

DIODE was modified by A. Ward from a gas ionization code to add modeling capability to bipolar semiconductor devices. The original gas code, which was used for high field breakdown studies in gases, was based on many of the same equations used in DIODE, but with different material parameters. For over a decade, DIODE has been successfully used by Ward in device modeling—and this is indicated by the numerous papers on semiconductor device physics listed in reference 1. A paper recently written by Ward, "Prediction of Second Breakdown Power," is to be published in the *Journal of Electrostatics*. It is based on modeling using DIODE and shows how certain device parameters obtained by low-level forward bias measurements can be useful in predicting reverse bias failure.

DIODE[1] solves the macroscopic drift-diffusion and thermal equations in one dimension. Although available, a two-dimensional version is very expensive to run and is less stable in numerical convergence than the one-dimensional version. Since the ratio of the junction diameter (80 μm) to the junction thickness (15 to 3 μm) of the test devices used in this study varies between 5 to 27, a one-dimensional solution was deemed satisfactory as long as the modeling does not include current filamentation and hot spot formation in the junction.

Our objective is to calculate the p-n junction current density versus voltage dependence from first breakdown to the negative differential resistance (NDR) regime, and to determine the latter by plotting several I versus V points and observing those values that lead to a negative slope. Since experimental pulse data taken by Tasca[15] and others show that the NDR is an unstable bias regime and that device failure will occur if no current limiting exists, the criterion for failure adopted in this study is that device degradation will occur after first breakdown in the NDR knee. This criterion is very convenient to invoke when it becomes necessary to compare computer modeling and experimental data (since modeling beyond the NDR knee is very difficult to implement), but it is necessary to realize that it ignores different levels of device degradation.

Ward has shown that the voltage and current density regimes in a p-n junction can be divided into four important bias regions between first breakdown and second breakdown:[1] the space-charge-free region, the space-charge-limited region, the double avalanche region, and the negative-differential-resistance region, as shown in figure 11.

The current density changes exponentially with voltage in the space-charge-free region, and follows a power law dependence in the space-charge-limited region. This departure from an exponential to power law dependence is due to space-charge resistance increase. This resistance becomes significant when there is an excess generation of electrons (due to impact ionization) and a saturated carrier velocity (limited by optical phonon interaction) causing an accumulation of negative charge in the depletion layer. This space-charge cloud is eventually neutralized by the presence of holes (leading to the NDR regime). The holes are also generated by impact ionization, but at approximately an order of magnitude higher electric field levels than for electrons. The cancellation of the electron space charge resistance is abrupt and results in a runaway current, leading to device failure. Knowing the critical time to reach the NDR regime is essential for protecting or hardening a semiconductor device against an electromagnetic transient.

Input Parameter Data for DIODE

DIODE requires that the p-n junction to be modeled be divided into a maximum of 100 cells. The anode starts at $x = 0$, while the cathode ends at $x = l$, l being the diode thickness. Within each of these intervals, the digitized doping concentration values are used for both the p and n regions. The boundary conditions for both the current density and the current continuity equations require that the initial electron current density be specified at the cathode $(x = l)$ and the initial hole current density be specified at the anode $(x = 0)$. The potential across the diode is given when one is solving Poisson's equation, with the rest of the junction's boundary conditions. This is where the bias (including any dc) is specified from the pulsed waveform. The input file data for DIODE consists of the silicon material constants, the transient waveform, the external circuit parameter values and the doping profile (shown in fig. 10).

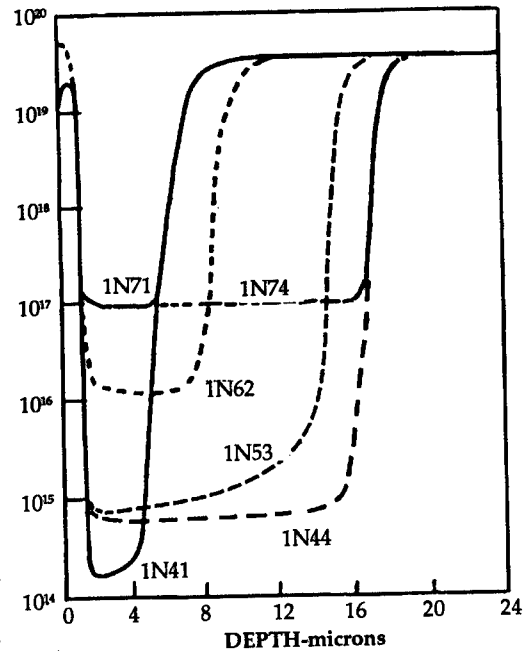The following is a list of the silicon material constants used in DIODE:

Figure 10. Epitaxial p-n junction profiles for lots 1N41, 1N44, 1N53, 1N62, 1N71, and 1N74.
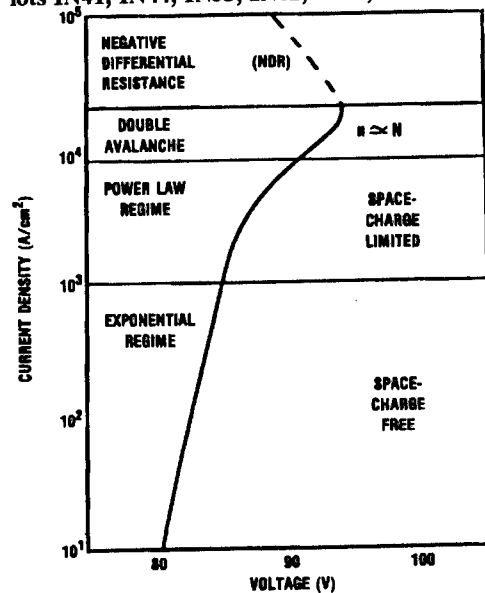
Figure 11. The four bias regions for a p-n junction between first and second breakdown.

Density	2.328 g/cm³	Energy gap (at 300 K)	1.12 eV
Specific heat	0.7 J/g °C	Intrinsic carrier concentration	$1.45 \; 10^{10}$ 1/cm³
Thermal conductivity (at 300 K)	1.5 W/cm °C	Melting point	1410 °C
Thermal diffusivity	0.9 cm²/s	Minority carrier lifetime	2.5 ms
Dielectric constant	11.8		

When one is modeling a p-n junction, DIODE requires that realistic external resistance, inductance, and capacitance values be included in the network shown in figure 12. This is to assure that the voltage and current of the junction reach values that are consistent with real circuit time constants. The resistor R is set at 50 Ω, which is the source impedance and the impedance of our measurement setup. The parasitic and device capacitance and inductance values of the wafer probe were measured at 2 pF and 10 nH, respectively, and are part of the external circuit parameters.

Computer Output Data

A series of computer calculations were made to obtain current density versus voltage profiles for the lots that exhibited current mode type failure and did not permit measurement of the NDR regime. These unstable junction lots are 1N41, 1N44, and 1N53.

For the lightly doped lots, the time step had to be set at about 50 fs and this value was reduced whenever any instability was noticed in the calculations. When further reduction of this time step did not reflect changes in the numerical output values, the previous time step value was kept to maintain the most economical run possible. The time step values are usually one to two orders of magnitude less than the dielectric relaxation time constants. From the four different doping densities the relaxation time constants ($\tau = \varepsilon/\sigma$) are computed as follows:

$$t_1 = 4 \times 10^{-11} \text{ s} \quad \text{for } N = 10^{14} \text{ cm}^{-3} \qquad t_3 = 4 \times 10^{-13} \text{ s} \quad \text{for } N = 10^{16} \text{ cm}^{-3}$$
$$t_2 = 4 \times 10^{-12} \text{ s} \quad \text{for } N = 10^{15} \text{ cm}^{-3} \qquad t_4 = 4 \times 10^{-14} \text{ s} \quad \text{for } N = 10^{17} \text{ cm}^{-3}$$

Modeling lots 1N41, 1N44, and 1N53 presented no major problems with numerical convergence of the electrical field in the junction, nor were the run times excessive. In all, 21 runs were needed to determine the NDR region for these three lots. These results are summarized in figure 13 with the experimental data.

Junction lots 1N41 and 1N44 showed no positive differential resistance (PDR) before entering the negative differential resistance regime (NDR). An increase in the epi thickness from 3 μm (1N41) to 15 μm (1N44) changed only the first breakdown voltage; it did not add any amount of PDR. Increasing the doping concentration from 10^{14} (1N41, 1N44) to 10^{15} (1N53) instead inserted a very weak amount of PDR before the NDR knee; however, the PDR proved to be too small to observe experimentally in a 50-ohm measurement system.

The experimental data showed that once the epitaxial doping level was increased to 10^{16} cm^{-3}, it was possible to obtain stable measurements of the NDR regime, for 30-ns pulses or less. However, at this doping level the cost of the computer runs became exces-
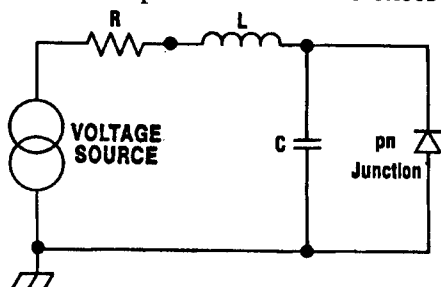

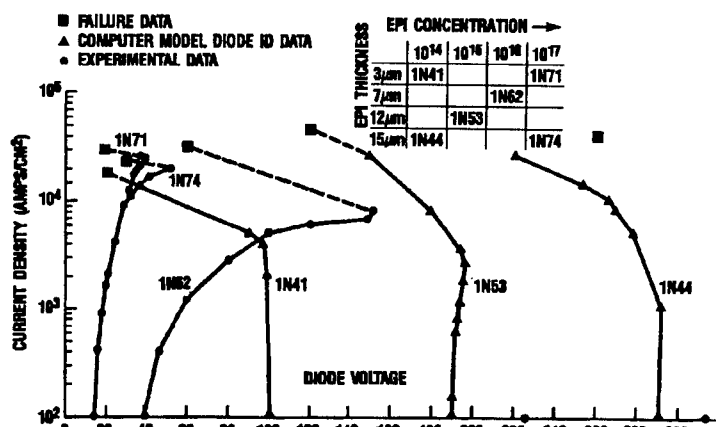Figure 12. DIODE's external RLC circuit.

Figure 13. Experimental data (1N71, 1N74, 1N62) and computer modeling (1N41, 1N44, 1N53, 1N62) of negative differential resistance regime for a wide range of p-n junction parameter variations between first breakdown and failure.

551

sive, since the time step had to be reduced by two orders of magnitude. Fortunately, the combination of experimental data with computer modeling together determined the NDR regime for all six lots.

Equipment Requirements

The circuit shown in figure 14 was designed to measure the dynamic current versus voltage relationship for the diode lots. It consists of a charged transmission line for generating a 1- to 30-ns video pulse. With this pulse-forming network, the generated pulse has a width that is twice the delay length of the charge line and half the charged voltage. The discharge switch selected is a mercury wetted, 300 lb/in.^{2}, H_2 atmosphere, magnetically activated mechanical switch. This type of high-voltage switch is capable of withstanding 6 to 8 kV and can switch states in only 350 ps. For a 1000-V pulse this represents a dV/dt of 3×10^{12} V/s. Most pulsed data previously taken had values less than 10^{10} V/s. This fast rate of rise in the voltage waveform permits one to overvoltage the junction in times less than the thermal time constant.

The two delay lines in the circuit are needed to center the voltage and current pulses with the center of the oscilloscope trace and to properly time the scope triggering circuit with the pulse.

Since the experimental test junctions were not packaged, the silicon wafers were probed on a Wentworth probe station. The bandwidth for a commercial probe station is usually very poor (less than 100 MHz) and cannot be used for these measurements, unless it is modified. The poor bandwidth is due to the improper ground return of the probe head. Introducing a low-inductance ground contact, in the shape of a loop, while the wafer was being probed, improved the bandwidth by over an order of magnitude to almost 2 GHz. Without this simple, yet effective, modification, executing the experimental part of this study would have been impossible.

The pulsed data for the wafer lots were displayed on Tektronix 7104 oscilloscopes and recorded on Polaroid film, while the very fast risetime data were recorded on a Tektronix 7250. The curve tracer data were taken after each pulse to determine if the device had failed, and when it was damaged, the data were recorded on film. For the experimental data, the criterion used for device degradation is a permanent decrease in the static breakdown voltage of 10 percent or more and an increase of an order of magnitude or more in the reverse leakage current. These criteria assure that the test devices had reached the negative differential resistance knee. For the lightly doped samples, 1N41, 1N43, and 1N44, failure occurred during the 30-ns pulses without any current plateau being exhibited, indicating current mode breakdown. For the heavier doped diodes, 1N71, 1N74, and 1N62, current mode breakdown was not observed, and failure occurred only for longer pulses, indicating the conventional energy-dependent second breakdown or thermal failure. It is important to note that these observations were made with a 50-ohm source.
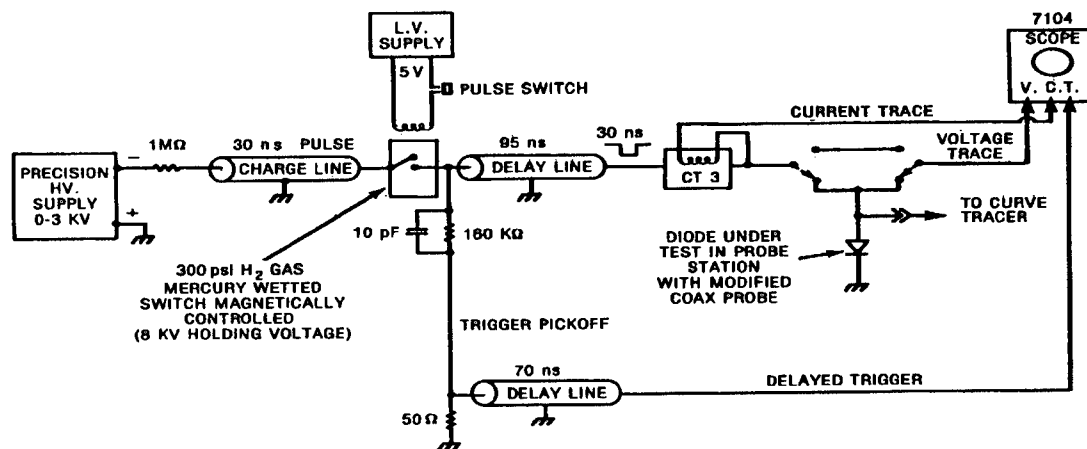

Figure 14. High amplitude pulse-forming network and p-n junction measurement setup.

Combined Experimental and Modeling Results

The combined experimental and computer modeling data representing the dynamic current density versus voltage for the six lots are given in figure 13. These data are plotted between first breakdown and failure. Because of the lack of any significant positive differential resistance before failure, the experimantal measurements of lots 1N41, 1N44, and 1N53 were very difficult to take with 30-ns pulses. It is believed that the lack of positive differential resistance before the negative differential resistance regime is responsible for the "current mode" failure instabilty. These experiments were repeated many times for different p-n junctions on the same wafer, with consistent results. Although the computer modeling showed a slight amount of positive differential resistance for lot 1N53 at 10^3 A/cm^2, this value was not sufficient to operate on a stable load line for a 50-ohm system. Hence these devices failed without an observable NDR knee. Even with the use of a Tektronix 7250 oscilloscope with 50-ps/division sweeptime (which represents at least an order of magntiude improvement over any previous measurement), no transition or current plateau could be observed.

The computer calculations show that the NDR regime (unstable operation) in the epi region occurs when the mobile carrier concentration exceeds the background doping concentration ($n = N$), indicated in figure 11. This finding is discussed by Ward[1] and in this study it occurs for four different values of the six lots. The current densities ($J = nqv$) at the point of instability are calculated to be as follows:

$$J(\text{at } 10^{14}) = (1.6)10^2 \text{ A/cm}^2 \quad \text{for 1N41, 1N44} \qquad J(\text{at } 10^{16}) = (1.6)10^4 \text{ A/cm}^2 \quad \text{for 1N62}$$
$$J(\text{at } 10^{15}) = (1.6)10^3 \text{ A/cm}^2 \quad \text{for 1N53} \qquad J(\text{at } 10^{17}) = (1.6)10^5 \text{ A/cm}^2 \quad \text{for 1N71, 1N74}$$

Since the velocity of the mobile carriers saturates at 10^7 cm/s, these current density values vary by three orders of magnitude. For a given pulse amplitude, a 50-ohm source limits the current magnitude. Thus in the lightly doped diodes, it is possible for currents to reach the NDR knee and to lead to device failure. That is, the electron avalanche process proceeds to double avalanche, and hence to excessive currents through filament formation. In contrast, for the more heavily doped diodes, the 50-ohm system limits current magnitude to values less than those listed above, restricting the avalanche process to the space charge limited region (fig. 11). Thus failure can occur only with longer duration pulses that result in thermal carrier generation, and ultimately, filamentation. It may be possible to obtain current mode breakdown in the heavier doped diodes by reducing source resistance to a few ohms or by applying much larger amplitude voltage pulses. However, such low resistances may be unrealistic in actual systems. Similarly, current mode breakdown in the lightly doped devices can be avoided by using high source resistance. This conclusion is supported by the shapes of the current-voltage curves shown in figure 13. For the lightly doped diodes, the slope of the current following first breakdown is nearly vertical (no positive differential resistance). Following first breakdown, current increases exponentially with voltage in the electron avalanche regime and, with no current limiting due to source resistance, proceeds through the space charge region and double avalanche to the NDR knee. The most heavily doped diodes exhibit a pronounced PDR region. Here with currents limited to the space-charge-limited region, heating of the diode occurs, resulting in first an increase in resistance (reduced mobility) and then a decrease in resistance (increase in intrinsic carrier concentration). Figure 3 indicates the static resistivity of silicon as a function of temperature.

Punch-Through Factor (PTF) and Breakdown

The PTF is the ratio of the distance over which the electric field extrapolates to zero (at first breakdown) to the physical width of the junction, as depicted in figure 15. The PTF can have values greater or less than one, depending on where the field extrapolates to zero in the junction. For lots 1N41 and 1N44, where the electric field is almost uniform across the junction, the PTFs are greater than one. Since the doping concentrations for these two lots are roughly equal, the difference in the first breakdown voltages is due largely to the change in the epi thickness.

The case for lots 1N71 and 1N74 is different. Although the epi thickness still varies from 3 to 15 µm, there is no difference between the first breakdown voltages. For these two lots the PTF is less than one, which suggests that most of the applied voltage is dropped across a fraction of the epitaxy layer. Calculations made by Ward[1] at 10^{17} cm^{-3} show that the field depletes in a few microns at first breakdown.

In figure 16 a wide range of epitaxial doping concentrations and epitaxial thickness parameters with related PTF values, as calculated by Ward,[1] are compared with the PTF values of the experimental junctions used in this study.

Direct Observation of Breakdown

As mentioned earlier, doping density and epi thickness determine first breakdown in silicon. Doping density, PTF, and junction area determine the current levels at which second breakdown occurs. Until this study was completed, it was not clear what role junction topology plays in device failure. Although it is not possible to model a multidimensional topology problem on DIODE, it has been possible to collect experimental data from odd-shaped junctions fabricated on the same test wafer, such as a triangular junction with area equal to the circular ones (5025 µm²). Experimental results showed that the change in topology affected only the first breakdown voltage value (since less voltage is needed to generate avalanche breakdown at the junction corners), but did not alter the failure level which remained the same as for the circular junctions. This shows that total area, and not topology, is important for predicting failure. Figure 17 shows the comparison of the I versus V characteristics for the two types of junction topologies at first breakdown and how they follow the same slope at higher bias levels. By detecting photons emitted by recombination currents, it was possible to identify the location of current hot-spots in the junction and observe where they first formed. At low bias levels they seemed to concentrate only at the perimeter. At higher levels they spread uniformly to the rest of the junction. Ironically, even at low current levels, computer modeling of the junction can be a multidimensional problem due to the nonuniformity of ionization impact formation throughout the junction. At higher bias levels where the current
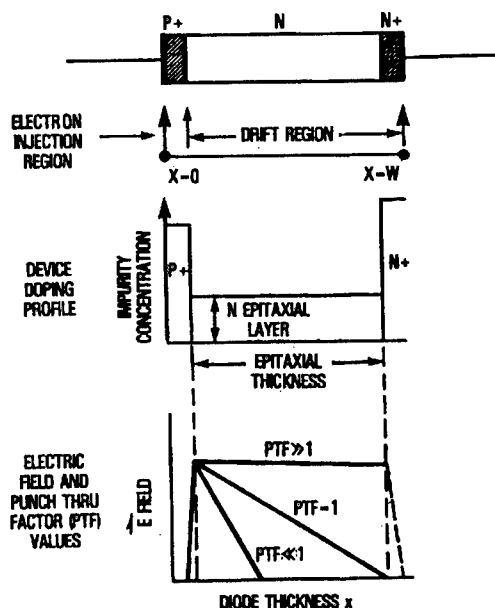

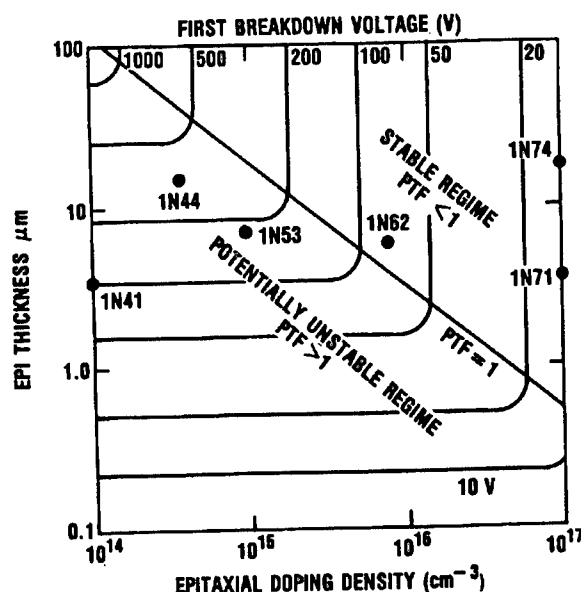Figure 15. Graphic representation of the punch-through factor (PTF) in a p-n junction.

Figure 16. PTF values for lots 1N41 through 1N74.

spreads more uniformly, the modeling problem simplifies to a one-dimensional analysis for circular junctions, before requiring again a multidimensional analysis when current filamentation occurs after the negative differential resistance regime. It was for this reason that circular junctions were chosen for the experimental devices. While integrated circuit technologies rely most often on a rectangular geometry (also known as "Manhattan Geometry") many discrete devices still use a circular topology.

Summary of Short-Pulse Effects

Junction diodes with a wide range of doping density and epitaxial thickness have been studied here, making use of new experimental techniques to determine the nature of current mode breakdown. Pulses with risetimes greater than 10^{11} V/s were used to study this fast phenomenon. This represents an improvement by nearly two orders of magnitude over previous work. Interpretation of the data and computer simulation leads to the following important results:

(1) Current mode type of failure is most likely to occur in lightly doped silicon junction technologies such as analog devices, PIN or avalanche diodes, rectifiers, and signal detectors. As discussed, over-voltaging these devices is easier to accomplish in circuits with source impedances of tens of ohms. In heavier doped devices, such as TTL digital devices, Zener devices, and voltage regulator diodes, or in parasitic devices associated with VLSI circuits, the conventional energy-dependent (thermal) breakdown is likely to occur unless the source impedance is extremely low. Identifying these technologies is important for successfully hardening against electromagnetic transients.

(2) A model for current mode breakdown has been presented. Essentially, if the resistance of the external circuit and the applied voltage permits sufficient current to flow so that the carrier concentration in the epitaxial region exceeds the background doping, breakdown can occur on electronic time scales. The avalanching of holes leads to a cancellation of the space charge resistance of the avalanching electrons, resulting in a negative differential resistance. For lightly doped devices, the required current levels can be modest. For heavily doped devices, the required current levels for purely electronic breakdown is often unattainable in typical circuits, and thus conventional thermal breakdown is observed since significant heating of the junction is the responsible failure trigger mechanism.

(3) Current mode and energy dependent breakdown have been associated with punch-through factor (PTF), and it was demonstrated
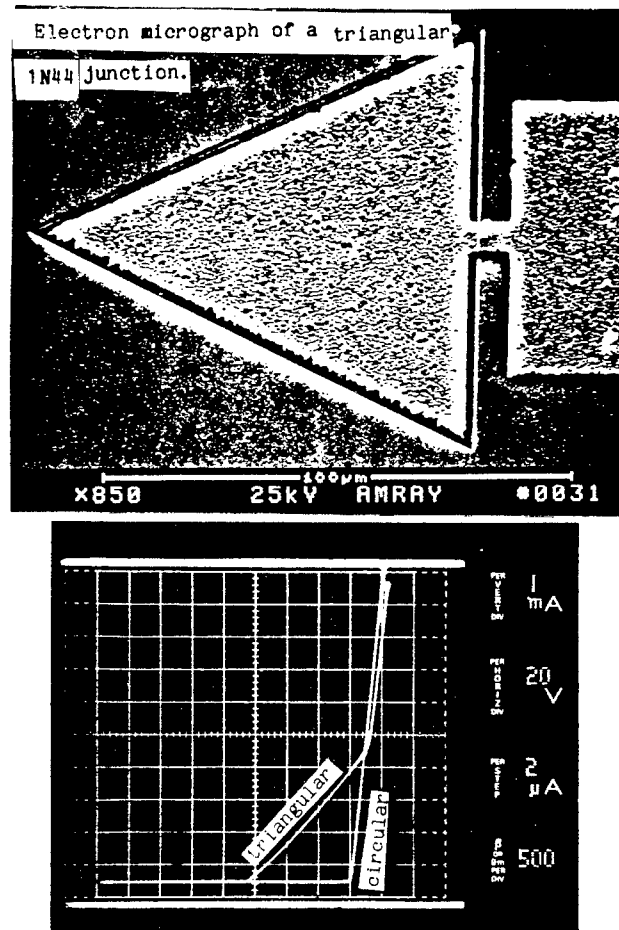

Figure 17. I versus V comparison for triangular and circular junction topologies.

555

that devices with a PTF greater than unity are susceptible to current mode failure (1N41, 1N44, and 1N53).

(4) First breakdown was directly observed in diodes of different topology. It was found that first breakdown depended on doping density, epitaxial thickness, and shape. In all cases studied here, first breakdown occurs at the edges of the device. However, topology had no effect on device failure for either current mode or thermal second breakdown.

Finally, concerning the issue of whether energy is the appropriate variable to use when one is addressing semiconductor failure, no general statement can be made. At one extreme the problem is similar to dielectric breakdown in insulators versus heat damage in metal conductors. Although a thermal process is responsible for the actual (observable) damage, the failure mechanism can be electronic or thermal. Failure energy is a meaningless parameter in the study of insulator breakdown. The fact remains that it is still impossible to measure currents in the picosecond time scale and for those devices that show current mode failure, it is not possible to accurately quantify any failure level except by expressing the amplitude and duration of the pulse causing the overvoltage. For those devices that show conventional thermal failure, energy remains an appropriate variable to use. This study distinguishes between these two failure groups and provides insight on their differences.

References

1. Alford L. Ward, *Calculations of Second Breakdown in Silicon Diodes*, Harry Diamond Laboratories, HDL-TR-1978 (August 1982).

2. J. Tauc and A. Abraham, Thermal Breakdown in Silicon p-n Junction, *Phys. Rev.*, Vol. **108**, pp. 936–937 (1957).

3. H. C. Chen, W. M. Portnoy, and D. K. Ferry, Doping Dependence of Second Breakdown in a p-n Junction, *Solid State Electron.*, Pergamon Press, Vol. **14**, pp. 747-751(1971).

4. D. C. Wunsch and R. R. Bell, Determination of Threshold Failure Levels of Semiconductor Diodes and Transistors Due to Pulse Voltages, *IEEE Trans. Nucl. Sci.*, **NS-15**, No. 6 (December 1968).

5. W. B. Smith, D. H. Pontius, and P. P. Budenstein, Second Breakdown and Damage in Junction Devices, *IEEE Trans. Electron Devices*, Vol. **ED-20**, p. 731 (1973).

6. R. J. Minniti, Investigation of Second Breakdown in Semiconductor Junction Devices, McDonnel Douglas Corporation, *Avionics Technical Note*, **ATN 73-002** (15 June 1973).

7. H. A. Schafft, Second-Breakdown—A Comprehensive Review, *Proc. IEEE*, Vol. **55**, p. 1272 (1967).

8. Daniel J. Fleming, Thermal Breakdown Delay Time in Silicon p-n Junctions, *IEEE Trans. Electron Devices*, Vol. **ED-18**, No. 2 (February 1971).

9. P. J. Kannam, F. G. Ernick, and J. Marino, Second Breakdown of Double Epitaxial Transistor, *IEEE Trans. Electron Devices*, Vol. **ED-15**, No. 2 (February 1968).

10. K. L. Chiang and P. O. Lauritzen, Thermal Instability in Very Small p-n Junctions, *IEEE Trans. Electron Devices*, Vol. **ED-17**, No. 9 (September 1970).

11. William M. Portnoy and Larry J. Scotts, *Second Breakdown Behavior of a Single Microplasma p-n Junction*, Harry Diamond Laboratories, HDL-CR-76-054-1 (August 1976).

12. J. Miletta, L. W. Ricketts, and J. E. Bridges, *EMP Radiation and Protective Techniques*, Wiley-Interscience Publication (1976).

13. W. J. Orvis, J. H. Yee, G. H. Khanaka, and D. L. Lair, *Theoretical Modeling of EMP Induced Breakdown Effects in Semiconductor Devices*, Lawrence Livermore Laboratory, UCID-19900 (June 1983).

14. H. B. Grutchfield and T. J. Moutoux, Current Mode Second Breakdown in Epitaxial Planar Transistors, *IEEE Trans. Electron Devices*, Vol. **ED-13**, No. 11 (November 1966).

15. D. Tasca, J. Peden, and J. Miletta, Non-destructive Screening for Thermal Second Breakdown, *IEEE Trans. Nucl. Sci.*, Vol. **NS-19**, No. 6 (December 1972).

Fluctuating Deep Level Trap Occupancy Model for 1/f and Low-Frequency Noise in Semiconductor Transistors

Patrick A. Folkes, Dr.
U.S. Army Electronics Technology and Devices Laboratory
LABCOM, Fort Monmouth NJ 07703-5000

INTRODUCTION

Despite decades of study, the detailed mechanisms of 1/f noise in semiconductor devices, defined as current fluctuations whose spectral density varies as 1/f over a large frequency range, are not well understood[1,2]. Several theoretical models for 1/f noise in semiconductor devices have been proposed[1-6]. However the basic characteristics of semiconductor 1/f noise, such as the spectral density and range and its temperature dependence, cannot be explained self-consistently by any of these theories. Recent observations of two-level[7] and low-frequency conductance fluctuations[8] in small-area devices have confirmed the importance of interface traps as a 1/f noise source; however the kinetics of charge transfer between the interface states and the bulk and its relationship to a 1/f noise spectrum is not understood[1,9,10]. Experimental results have confirmed the fact that bulk traps generate low-frequency and 1/f noise in GaAs field-effect transistors(FETs)[11-14] and in AlGaAS/GaAs high-electron-mobility transistors(HEMTs)[15,16]. Recent measurements show that the 1/f noise in GaAs FETs consists of a surface component and a bulk component which can dominate the surface component over a wide range of applied voltages. Theoretical[17] and experimental studies[18,19] of generation-recombination(g-r) noise in FETs show that the dominant mechanism for this type of low-frequency noise, which is heretofore associated with a single time constant, is the fluctuating occupancy of deep-level traps in the depletion region. The dependence of the relaxation time on the free electron density in the depletion region was neglected in reference 17 and the relationship between 1/f and g-r noise was not explored.

Recently we reported[20] the development of a self-consistent theory for 1/f noise generation by bulk traps which are located in the depletion region and which are arbitrarily distributed over energy relative to the flat-band Fermi level. Our results show, for the first time, that the resultant noise spectrum exhibits a 1/f dependence over a large frequency range if the energy distribution of the bulk traps satisfies a simple condition. In this paper we present a comprehensive review and application of the model to an FET biased over a wide voltage range. To my knowledge

this is the first work which quantitatively relates the spectral density and range and the temperature dependence of 1/f noise to the energy distribution of bulk traps and explains the experimentally observed dependence of the 1/f noise spectral density on electron mobility and applied voltage. We would like to point out that, although the model presented here is specific to the FET(or HEMT) structure, the underlying physical mechanism is applicable to other semiconductor devices and transport effects which are sensitive to fluctuations in the width of a depletion layer or in the number of carriers.
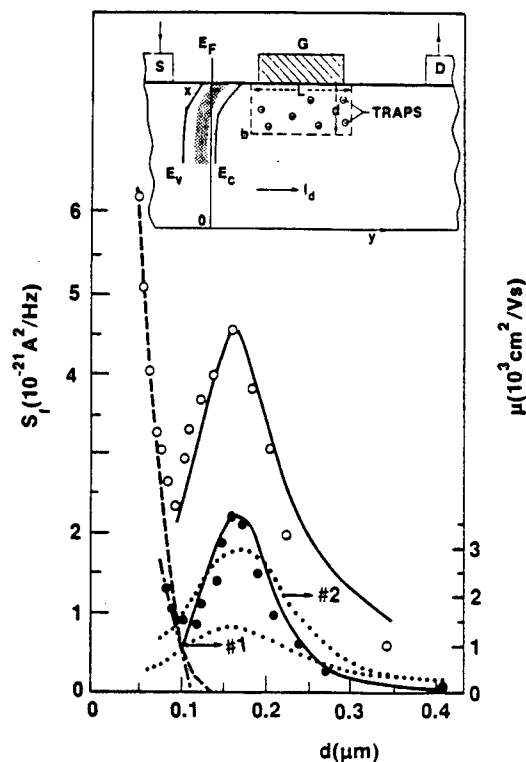

Fig.1- Experimental and theoretical fit of $S_f(0.1MHz)$ and $\mu(d)$ vs. depletion depth, V_D -0.08V.The insert shows the FET geometry.

THEORY
A. Low Frequency noise generated by single-energy bulk traps

The insert in Figure 1 shows the Schottky barrier FET conducting channel geometry for the case where the drain-source voltage V_D, is small compared to the voltage needed to induce velocity saturation and the gate length is long compared to the depletion depth. The band-bending in the depletion layer is also shown in Figure 1. In this low-field regime, the

lower boundary of the depletion layer at $x=b(V_g)$ remains approximately parallel to the surface of the gate for all gate voltages. The low-field analysis elucidates the physics involved in the capture and release of carriers in the depletion region. We will extend the analysis to the case where V_D is large in a later section. For simplicity, two-dimensional effects near the edge of the gate are ignored. The model is based on the following simple physical mechanism: random fluctuations in the occupancy of deep-level traps in the depletion layer induce charge fluctuations on the gate which result in fluctuations in channel resistance. This is precisely the same mechanism proposed by Sah to explain g-r noise[17]. In this paper we explicitly take into acccount the non-negligible free electron density and the variation of trap occupancy in the depletion layer and demonstrate that these effects significantly impact the low-fequency noise spectrum of deep-level traps.

In any semiconductor, fluctuations in the number of trapped electrons occur as a result of random electron and hole capture and emission processes. Considering the case of single-level acceptor centers in the depletion region, Sah showed that an initial trapped electron density fluctuation δn_{t0}, decays exponentially with the time constant τ, given by [17]

$$\tau(x)=1/[c_p p_1 + c_n(n_o(x)+n_1)] \tag{1}$$

where $c_n = \sigma_n v_{tn}$ and $c_p = \sigma_p v_{tp}$ are the electron and hole capture probabilities respectively, σ_n, v_{tn}, σ_p, v_{tp} are the capture cross-sections and mean thermal velocities for the electron and hole respectively and n_o is the steady-state free electron density in depletion region. In deriving (1), which is valid in the limit $\tau \gg \tau_e$, the transit time across the depletion region, it is assumed that σ_n and σ_p are field-independent. The quantities, p_1 and n_1 are given by

$$p_1 = n_i \exp((E_i - E_t)/kT) \tag{2}$$

$$n_1 = n_i \exp((E_t - E_i)/kT) \tag{3}$$

n_i is the intrinsic carrier density, E_i is the intrinsic Fermi level and E_t is the trap energy level. The free electron density in the depletion layer is given by the Boltzman distribution

$$n_o(x) = n_c \exp(e\phi(x)/kT) \tag{4}$$

n_c is the free electron density in the conducting channel at the edge of the depletion region and $\phi(x)$ is the electrostatic potential. A quasi-Fermi level for the electrons F_n, can be defined such that the trap occupancy f_t, is given by Fermi-Dirac statistics[21,22].

Due to its dependence on $n_o(x)$ the time constant varies smoothly

over a large range. For example, using the following GaAs material parameters[23,24] at $T = 300K$; $n_c = 10^{17}/cm^3$, $n_i = 1.79 \times 10^6/cm^3$, $\sigma_n = 8 \times 10^{-16} cm^2$, $\sigma_p = 4.3 \times 10^{-16} cm^2$, $v_{tn} = 4.4 \times 10^7 cm/s$, $v_{tp} = 1.6 \times 10^7 cm/s$, $E_t \approx 0.7eV$ we obtain $c_n = 3.5 \times 10^{-8}$ and $c_p = 6.4 \times 10^{-9} cm^3/s$. Equation (1) shows that the time constant varies over the range $2 \times 10^{-9} s \leq \tau \leq 15s$. A section of the depletion volume, where $c_n n_o(x) \ll c_n n_1 + c_p p_1$ may exist. In this region, for a given E_t, τ is constant and equal to its maximum value .

For an arbitrary free electron density profile in the conducting channel $n_c(x)$, it can be shown that

$$\frac{dG}{dQ} = \mu(d)/L^2 \tag{5}$$

where G is the channel conductance, $\mu(d)$ is the low-field electron mobility at the depletion layer boundary, L and Q are the gate length and charge respectively. For an elemental volume δv, located within the depletion layer at a distance x the charge fluctuation δQ, induced by the fluctuation in trapped electron density δn_t is $e \delta n_t \delta v$. Using (5) it can be shown that the current fluctuation arising from traps in δv, which have the same relaxation time $\tau(x)$, is given by

$$\langle (\delta I)^2 \rangle = V_D^2 (\mu(d)^2 e^2/L^4) \langle (\delta n_t \delta v)^2 \rangle \tag{6}$$

Following Sah's treatment we calculate the spectral density δS_f associated with fluctuations in δv using the Wiener-Khintchine theorem and the exponential dependence of δn_t and obtain

$$\delta S_f = 4V_D^2 e^2 \mu(d)^2 (W/L^3) N_t f_t (1-f_t) \tau(x)/(1+\omega^2 \tau(x)^2) dx \tag{7}$$

Where N_t is the density of deep-level traps, $\delta v = LWdx$ and $\omega = 2\pi f$. Since the low frequency fluctuations δn_t from different elemental regions are uncorrelated[21], the total low-frequency noise is obtained by integrating over the depletion region taking into account the spatial dependence of the Fermi-Dirac fluctuation factor $f_t(1-f_t)$ due to band-bending in the depletion layer. Due to band-bending a deep level E_t which is below F_n outside the depletion region, will cross F_n at some point in the depletion layer (see Fig. 2a). The fluctuation factor $f_t(1-f_t)$ is a sharply peaked function of energy with a full-width-half-maximum $\approx 4kT$ about its maximum value 0.25 at $E_t(x) = F_n$ as shown in Fig. 2b. Therefore, even if the deep level monoenergetic traps have a broad spatial distribution over the depletion region, they will generate low-frequency noise only from the relatively narrow region defined by the range of $E_t(x)$ over which $f_t(1-f_t)$ extends. For GaAs and most levels in Silicon , in the region where $E_t = F_n$, $n_0 \gg n_1 \gg p_1$, therefore $n_0 \approx (c_n \tau)^{-1}$. Using (4) it can be shown that
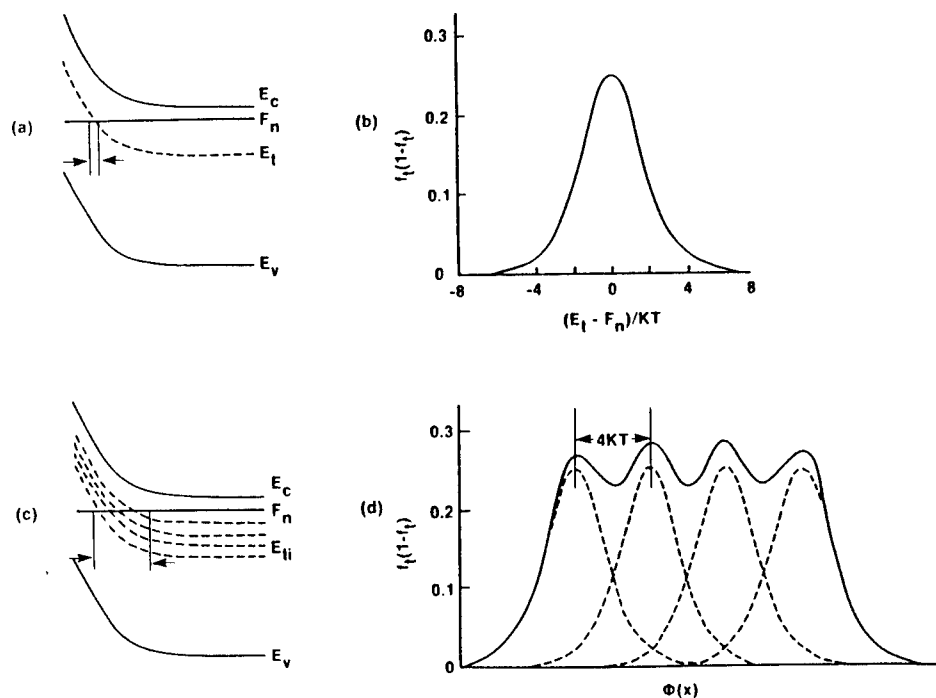
Fig.2- Energy band diagram and the associated fluctuation factor for (a) monoenergetic and (c)distributed energy traps.

$$f_t(1-f_t) = \frac{gc_n n_c \tau \exp(-(F_n - E_t)/kT)}{(1 + gc_n n_c \tau \exp(-(F_n - E_t/kT)))^2} \qquad (8)$$

where g=1 if E_t is nondegenerate and g=1/2 if E_t is spin degenerate.

The integration of δS_f over the depletion region can be transformed to an integration over τ . The spectral density of the low-frequency noise generated by the monoenergetic traps is given by

$$S_f(\omega) = 4(V_D \mu(d)e)^2(W/L^3)N_t \int_{\tau_m}^{\tau_1} f_t(1-f_t) \frac{\tau}{1+\omega^2\tau^2} \frac{dx}{d\tau} d\tau \qquad (9)$$

where τ_m and τ_1 are the lower and upper limits of the range of τ over which $f_t(1-f_t)$ extends.For numerical integration purposes τ_m and τ_1 are

563

defined as the lower and upper values of τ at $f_t(1-f_t)=2.5 \times 10^{-6}$, respectively. N_t is assumed to be constant throughout the depletion layer. Numerical calculations[25] have shown that, outside the $4\lambda_D$-wide transition region at the depletion layer boundary, $\phi(x)$ is given by

$$\phi(x) = \phi_0 (1 - x/d)^2$$

(10)

where ϕ_0 is the surface potential relative to the bulk, d is the depletion layer thickness calculated using the depletion approximation($n_0=0$ for $x<$ d) and λ_D is the Debye length. Combining (1), (7) and (10) and taking the derivative we obtain

$$\frac{dx}{d\tau} = \frac{\lambda_D}{\tau}[\ln(n_0/n_c)^2]^{-1/2} \qquad n_c \neq n_0$$

(11)

As noted above (11) is not valid in the depletion region boundary because the depletion approximation breaks down in this region and (10) is no longer valid. Substituting (11) in (9) and treating the factor $(\ln(n_0/n_c)^2)^{-1/2}$ as a constant since it is a slow-changing function of τ with an average value≈ 0.25, we obtain

$$S_f(\omega) = (V_D\mu(d)e)^2 (W\lambda_D/L^3)N_t \int_{\tau_m}^{\tau_1} \frac{c_n n_c \tau \exp(-(F_n - E_t)/kT) \, d\tau}{(1 + c_n n_c \tau \exp(-(F_n - E_t)/kT))^2 (1 + \omega^2\tau^2)}$$

(12)

Previous work on the g-r noise generated by monoenergetic traps[17] neglected the effect of the local free electron density and predicted a g-r spectral density S_L given by

$$S_L(\omega) = 0.5 (V_D\mu(d)e)^2 (W\lambda_D/L^3)N_t \tau_t/(1 + \omega^2\tau_t^2)$$

(13)

where τ_t is the relaxation time at the point where $E_t=F_n$. For comparison S_f and S_L are computed and plotted in Fig. 3 for the case where E_t is nondegenerate, $F_n-E_t =0.2eV$ in the flatband region, $n_c=10^{17}cm^{-3}$ and T=300K. Device size and other parameters are arbitrary. S_L displays the characteristic Lorentzian behavior: S_L is constant for $f \ll (2\pi\tau_t)^{-1} = 1.9 \times 10^5$ Hz and varies as $1/f^2$ for $f \gg 1.9 \times 10^5$ Hz. In contrast, S_f is constant for f < 10Hz and makes a relatively broad transition to a $1/f^2$ dependence for f > 5×10^7 Hz. Fig.3 shows that S_f varies approximately as $1/f$ over a narrow range of frequencies centered around $f= (2\pi\tau_t)^{-1}$. We define f_L and f_m as the frequencies at which S_f varies as $f^{-0.6}$ and as $f^{-1.4}$ respectively. Numerical evaluation of (12) predicts that $f_m/f_L \approx 72 = 10^{1.86}$. This ratio is independent of the trap energy and lattice temperature. In con-

564

trast if we apply the same definition of f_m and f_L to S_L we obtain $f_m/f_L \approx 4$. So the ratio f_m/f_L can be regarded as a broadening parameter.
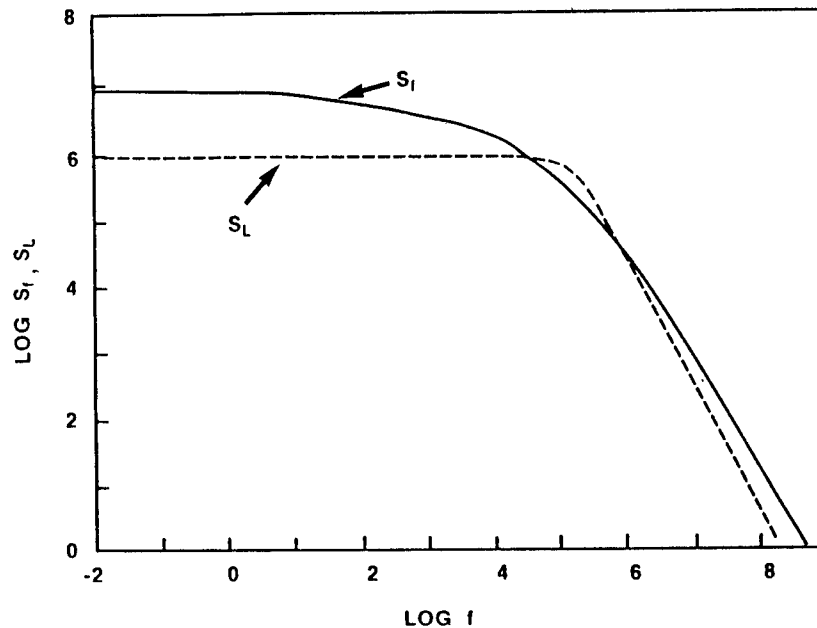

Fig.3- Theoretical low frequency noise spectral densities, S_f and S_L, for the traps described in Fig.2a.

B. 1/f and low-frequency noise generated by bulk traps with a discrete energy distribution.

Semiconductor devices under normal growth and processing conditions typically contain bulk and surface traps with a broad discrete distribution in energy. For example, fifteen different electron traps with activation energies in the range 0.2 - 0.83 eV, have been identified in GaAs[26]. A GaAs FET typically exhibits a subset of these traps which is characteristic of the particular growth and processing techniques used to produce the device. The spectral density of the the low-frequency noise generated by independent traps which are discretely distributed among n levels E_{ti} (i= 1 to n) is simply obtained by superposition,

$$S_f(\omega) = \sum_i S_{fi}(\omega) \tag{14}$$

where $S_{fi}(\omega)$ is the spectral density generated by traps with the energy E_{ti} given by (12). Note that only traps which are below the Fermi level in the flat-band region will generate low-frequency noise. From (14) we can

define an effective fluctuation factor as $\Sigma N_{ti} f_{ti}(1-f_{ti})$, where N_{ti} and f_{ti} are the density and occupancy of a particular level respectively. In Figs. 2c and 2d we show the band-bending and the effective fluctuation factor respectively for traps which are equally distributed among the four levels defined by F_n-E_t =0.1, 0.2, 0.3 and 0.4 eV at 300K. Note that the spacing between levels is 4kT. Fig. 2d shows that the effective fluctuation factor of these levels at 300K has a full-width-half-maximum $\approx$ 0.4eV and could be reasonably approximated by a single broad pulse. In general if the trap energy distribution is such that the separation between adjacent levels $E_t \leq$ 5kT for some range of T then the effective fluctuation factor will be broadened over this range of temperature. In this sense we say that these levels are coupled. Using (12) and (14) we numerically computed $S_f(\omega)$ for the traps which are shown in Fig. 2c with $n_c = 10^{17} cm^{-3}$ and T= 300K. The trap density is assumed constant in the depletion region. The other parameters in (12) are not specified here and S_f is plotted in arbitrary units in Fig. 4.
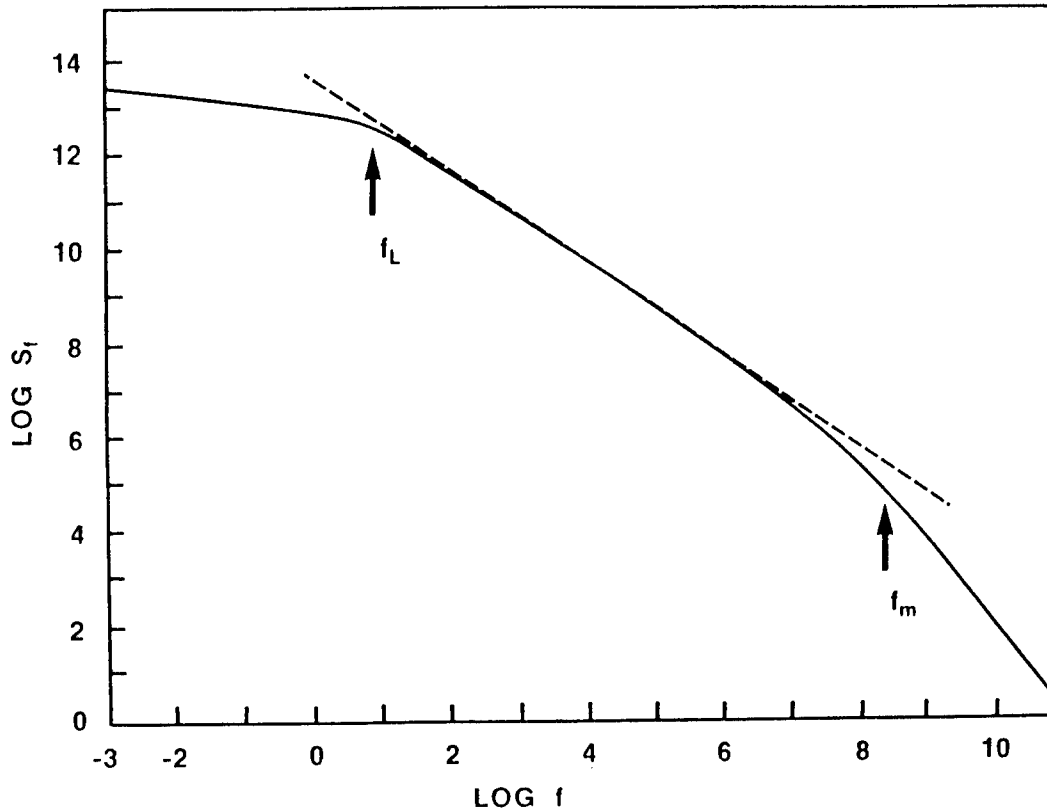

Fig.4- Calculated 1/f noise spectral density for the traps described by Fig.2c.

S_f displays a 1/f dependence over the large frequency range

30 Hz $\leq$ f $\leq$ 1.5x10^{7}Hz. S_f is almost frequency independent for $f < 0.1$Hz and varies as $1/f^2$ for $f \geq 10^9$Hz. In this case $f_m/f_L = 10^{7.4} = (10^{1.86})^4$. By extrapolation, if the traps are distributed among n coupled levels then the resultant spectral density will vary as $1/f$ over a maximum range of frequencies which can be characterised by $f_m/f_L = 10^{1.86n}$.

From (12) and (14) it is clear that the low-frequency noise spectrum of bulk traps is sensitive to their distribution over energy relative to the flat-band Fermi level. The low-frequency noise generated by bulk traps with an arbitrary distribution over n energy levels will exhibit a $1/f$ dependence only if the trap distribution over energy is such that $\Sigma N_{ti} f_{ti}(1-f_{ti})$ is a broad pulse whose full-width-half-maximum $\approx 4nkT$. The spectral range is determined primarily by the deepest and highest trap energy levels. An approximate analytical expression for the $1/f$ noise generated by a set of coupled levels can be obtained by approximating the effective fluctuation factor (see Fig.2d) as a square pulse. We set $f_t(1-f_t) = 0.25$ in the range $\tau_M \leq \tau \leq \tau_S$ and $f_t(1-f_t) = 0$ for τ outside this range. In this approximation τ_S and τ_M are the relaxation times at the half-maximum point of the effective fluctuation factor. Using this approximation and (11) in (9) we obtain

$$S_f(\omega) = 0.25 \cdot \mu(d)^2 e^2 V_D^2 N_t (W\lambda_D/L^3)(1/\omega)(\tan^{-1}\omega\tau_L - \tan^{-1}\omega\tau_S) \quad (15)$$

The spectral density given by (15) exhibits $1/f$ dependence for $\tau_S^{-1} \leq \omega < \tau_M^{-1}$. $S_f(\omega)$ is constant for $\omega \ll \tau_S^{-1}$ and varies as $1/f^2$ for $\omega \gg \tau_M^{-1}$. Experimental measurements of the $1/f$ noise spectrum in FETs typically can only access the frequency range where $\tau_S^{-1} < \omega < \tau_M^{-1}$, that is where S_f has a pure $1/f$ dependence. In this range (15) can be reduced to

$$S_f(\omega) = (\pi/8)\, \mu(d)^2\, e^2\, V_D^2\, N_t(W\lambda_D/L^3)\,(1/\omega) \quad (16)$$

COMPARISON WITH MEASUREMENTS ON FETs IN THE LOW-FIELD REGIME

Measurements on gold-doped silicon FETs[18] show that the observed low-frequency noise spectrum generated by the monoenergetic gold traps agrees with our theoretical results. From the data presented we determine that $f_m/f_L = 68$ in good agreement with our theoretically predicted value of $f_m/f_L = 72$. Fig.1 shows (with dots) measurements of the $1/f$ noise spectral density at f= .1MHz[11] and the low-field electron drift mobility of two GaAs FETs as function of depletion depth. Both devices are of similar design but device #1 used an ion implantation process to create its conducting channel. $n_c = 10^{17}$cm^{-3}, L=1μm and $V_D = .08$V for both devices. Fig. 1 shows that the $1/f$ noise consists of a bulk component and a surface compo-

nent, which by extrapolation decreases to zero at d= 0.12μm. It is clear that the bulk 1/f noise is correlated with $\mu(d)$. Equation (16) predicts that $S_f \propto \mu(d)^2$. In Fig. 5 we show plots of the experimental bulk $S_f(0.1$ MHz) versus μ^2 for the two devices. The excellent straightline fit provides experimental confirmation of our theoretical model. We can determine N_t, the only free parameter, by fitting the measured bulk S_f - d profile with (16) using the measured $\mu(d)$ profile. The solid lines in Fig.1 show the calculated $S_f(0.1$MHz) as a function of d for the devices using $N_t= 7.8$ x10^{14}cm^{-3} for device #1 and $N_t = 6.9$x10^{13}cm^{-3} for device #2. For d >0.12μm there is excellent agreement with the measured bulk S_f - d profile; for d < 0.12μm the disagreement is due to the superposition of the surface 1/f noise component.
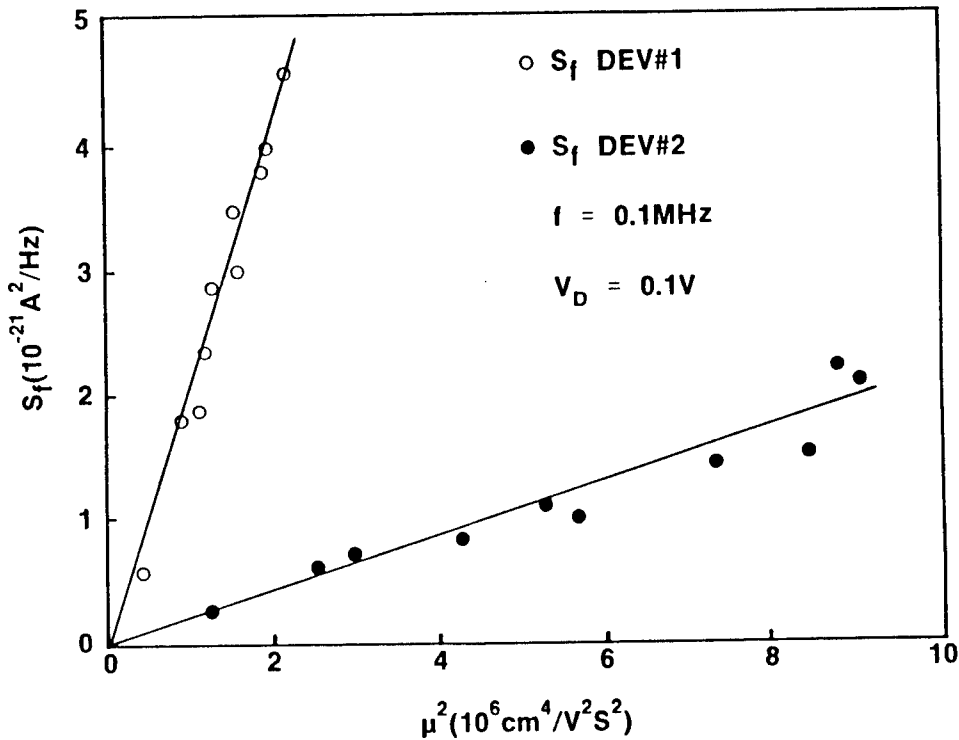

Fig.5- Experimental $S_f(0.1$MHz) vs. $\mu(d)^2$

VOLTAGE DEPENDENCE OF FET 1/f NOISE -THEORY AND EXPERIMENT.

For small drain-source voltage V_D, (16) shows that S_f is proportional to V_D^2. As V_D increases the depletion depth varies along the y-axis (see Fig.1). In essence this means that the channel resistance/cm varies along the channel. The fluctuating trap occupancy model still applies for the case where V_D is large but the nonuniform channel resistance results in a more complex voltage dependence. A detailed discussion is outside the scope of this paper but the primary results of our analysis are discussed
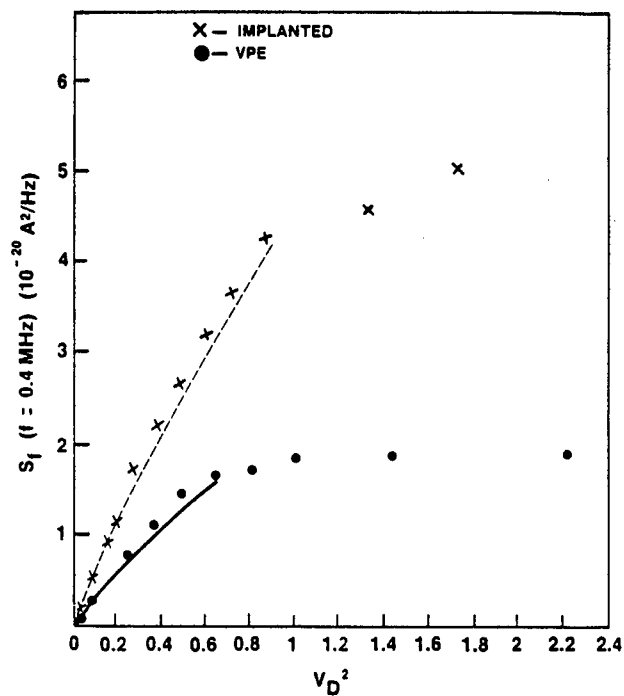
Fig.6- Theoretical and experimental drain voltage dependence of S_f with $V_g=0$.
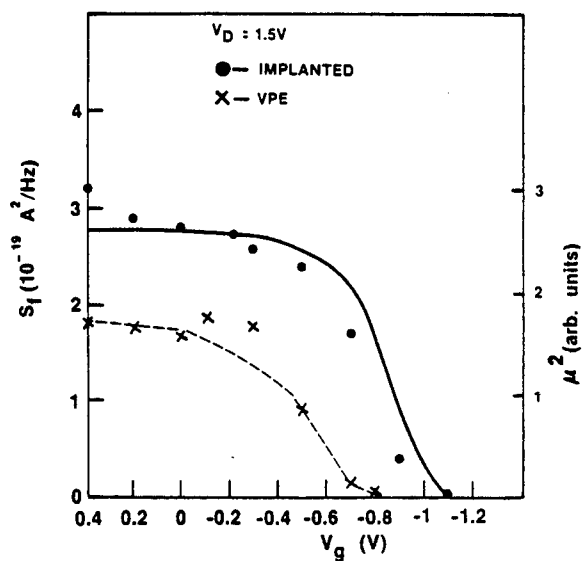

Fig.7- Measured S_f and average channel mobility vs. V_g with $V_D= 1.5V$.

below. Fig. 6 shows measurements(dots and crosses) and theoretical plots of S_f(0.4MHz) as a function of V_D^2 for the two FETs discussed above. Gate voltage V_g=0. The measurements show that $S_f \propto V_D^2$ for small V_D. For $V_D >$ 0.3 volt S_f increases sublinearly with V_D^2 and then saturates when $V_D > V_S$, the voltage needed to achieve velocity saturation. Fig. 6 shows that there is excellent agreement between theory and experiment for $V_D < V_S$. Our theory uses the gradual channel approximation for $V_D < V_S$. For $V_D > V_S$ our theory (not shown in Fig. 6) predicts that S_f saturates with a slope which is given by the output conductance in agreement with our experimental results.

When the FET is biased into saturation Fig. 6 shows that a very good approximate value for S_f can be obtained simply by assuming that S_f has a linear dependence up to V_S and completely saturates for $V_D > V_S$. Using (16) $S_f(\omega)$ is given by

$$S_f(\omega) = (\pi/8)\mu^2 e^2 V_S^2 N_t (W\lambda_D/L^3)(1/\omega) \tag{17}$$

μ is the average low-field electron mobility in the channel. Note that S_f implicitly depends on V_g because μ is a function of depletion depth. Fig. 7 shows(with dots and crosses) the measured dependence of S_f on V_g with V_D= 1.5V. The data shows that for the saturated FET S_f is insensitive to V_g over a wide range and then decreases sharply as the depletion boundary approaches the buffer layer interface. This is in sharp contrast with the low-field behavior shown in Fig.1. The lines in Fig.7 show the measured normalized square of the average electron mobility in the channel as a function of gate voltage. The data confirms that S_f/μ^2 is independent of gate voltage and provides further confirmation of our theoretical model. Equation (17) should be very useful for calculations of the 1/f noise amplitude in FETs.

The temperature dependence of FET 1/f noise can be obtained from (12). Our numerical calculations show that S_f is insensitive to temperature over a wide temperature range. Below some critical temperature S_f at a fixed frequency will switch from a 1/f to a $1/f^2$ dependence and will decrease sharply with a further decrease in temperature as shown by the theoretical plots in Fig. 8. This agrees well with experimental results on AlGaAs/GaAs Hemts[16].

SUMMARY

A quantitative theoretical model for 1/f and generic low-frequency noise in semiconductor transistors has been developed. The theory gives new insight into the fundamental mechanism of 1/f noise and explicitly accounts for many of the heretofore unexplained characteristics of 1/f and low-fequency noise in semiconductor devices. These results should be of considerable value to system and design engineers who are interested in calculating the low-frequency noise performance of transistors over a wide range of temperature and operating voltage.
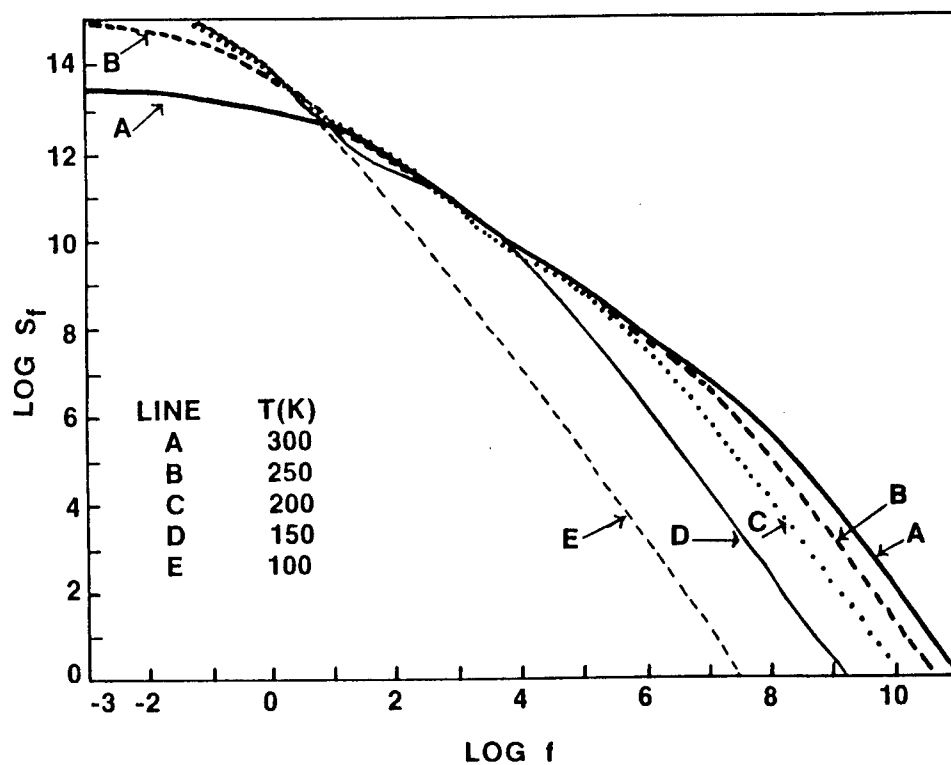
Fig.8- Theoretical temperature dependence of the 1/f noise spectral
density for the traps described by Fig.2c.

REFERENCES
1 M.B. Weissman, Rev. Mod. Phy., 60, 537, (1988).
2 P. Dutta and P.M. Horn, Rev. Mod. Phys., 53, 497, (1981).
3 A. Van der Ziel, Adv. Electronics and Electron Phys., 49, 225 (1979).
4 A.L. McWhorter, M.I.T. Lincoln Lab. Rept., No. 80, (1955).
5 F.N. Hooge Phys. Lett., A29, 139 (1969).
6 P.H. Handel, Phys. Rev. A, 22, 745 (1980).
7 K.S. Ralls et al, Phys. Rev. Lett.,52, 228 (1984).
8 C.T. Rogers and R.A. Buhrman, Phys. Rev. Lett., 53, 1272 (1984).
9 C.T. Rogers and R.A. Buhrman, Phys. Rev. Lett., 55, 859 (1985).
10 K.R. Farmer, C.T. Rogers and R.A. Buhrman Phys. Rev. Lett. 58, 2255 (1987)
11 P.A. Folkes, Appl. Phys. Lett., 48, 344 (1986).
12 C. Su, H. Rohdin and C. Stolte, Internat. Electron Dev. Meet. Tech. Digest, 601 (1983)
13 B. Hughes, N.G. Fernandez and J.M. Gladstone, IEEE Trans. Elect. DEv., ED-34, 733 (1987)
14 P. Canfield and L. Forbes, IEEE Trans. Elect. Dev., ED-33, 925 (1986).
15 L. Loreck et al IEEE Elect. Dev. Lett., EDL-5, 9 (1984).
16 S. Kugler, IEEE Trans. Elect. Dev., 35, 623 (1988).
17 C.T. Sah, Proceed. IEEE, 52, 795 (1964).
18 P.O. Lauritzen, Solid State Electron., 8, 41 (1965).
19 J. Graffeuil, K. Tantrarongroj and J.F. Sautereau, Solid State Electron., 25, 367 (1982).
20 P.A. Folkes, Appl. Phys. Lett., 55, 2217 (1989).
21 M. Lax, Rev. Mod. Phys., 32, 25 (1960).
22 W. Shockley and W.T. Read Jr., Phys. Rev., 87, 835 (1952).
23 J.S. Blakemore, J. Appl. Phys., 53, R123 (1982).
24 S.M. Sze, Physics of Semiconductor Devices, 2nd. ed. (Wiley, NY 1981), p. 320.
25 E. Wasserstrom and J. McKenna, Bell System Tech. J., 49, 853 (1970).
26 G.M. Martin, A. Mitonneau and A. Mircea, Electron. Lett., 13, 191 (1977).
27 P.A. Folkes, Appl. Phys. Lett., 48, 431 (1986).

Experimental and Computational Study of Collisions of Highly Excited Oxygen Atoms

Brad E. Forch, Dr.
Paul J. Dagdigian, Dr.
Millard H. Alexander, Dr.
*Andrzej W. Miziolek, Dr.
Ballistic Research Laboratory, SLCBR-IB-I,
Aberdeen Proving Ground, MD 21005-5066
The Johns Hopkins University, Department of Chemistry,
Baltimore, MD 21218
University of Maryland, Department of Chemistry,
College Park, MD 20742

1. INTRODUCTION: There is a considerable need for well-characterized diagnostic techniques for the quantitative measurement of the concentrations of species in flames and other combustion environments. The use of multiphoton excitation schemes for the detection of light atoms, such as oxygen, has drawn considerable interest in recent years.[1,4-14] Two different approaches have been employed in these methods, namely observation of the fluorescence emission from the multiphoton-excited resonant state[1,4-10] or absorption of an additional photon to produce ionization.[10-14] Both of these schemes have been utilized for the detection of hydrogen and oxygen atoms in flames. The former method has advantages over the latter in that it does not require the insertion of a probe into the medium and can also be used to obtain two-dimensional images of atomic concentrations in flames, as has recently been demonstrated.[15,16] It is also easier to discriminate against background photon emission than ionization.

Figure 1 illustrates the scheme which has been used for the detection of oxygen atoms.[1,4,6,8,10,15,16] Here, oxygen atoms are excited in a two-photon transition to the 3p ^{3}P state with 225.6 nm laser radiation and are detected by observation of the fluorescent emission to the 3s ^{3}S state at 844.7 nm. A number of parameters are required for the use of this diagnostic tool in a quantitative fashion. Recently, Bamford, et al.,[10] have reported absolute cross sections for the two-photon absorption process, as well as for one-photon photoionization of the ^{3}P state. Bimolecular collision quenching rate constants for this state have also been measured by several

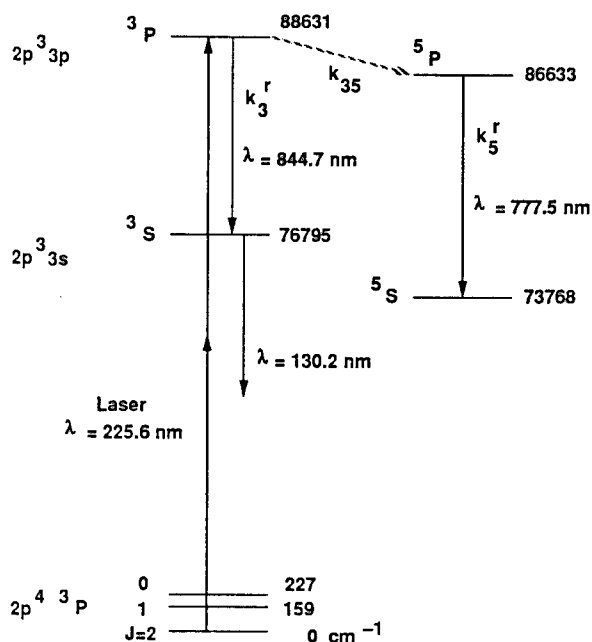groups.[1,4,10] These data are required to account for collisional effects in finite-pressure environments.

Figure 1. Relevant energy levels for two-photon excitation of oxygen atoms and excited-state collisional energy transfer. The electronic energies are given in cm^{-1}, and allowed radiative transitions are denoted, with wavelengths given in nanometers. The denoted rediative and collisional rate constants are defined in Section 3.2.

Miziolek and DeWilde[1] have also shown the potential importance of collisional excitation transfer processes in the collisional removal of this highly excited atomic state. They excited oxygen atoms by two-photon absorption at 225.6 nm in an atmosphere-pressure $CH_4-N_2O-N_2$ flame and observed within and through the flame front not only $^3P \rightarrow ^3S$ emission at 844.7 nm, but also $^5P \rightarrow ^5S$ emission of even greater intensity at 777.5 nm. Similar detection of 777.5 nm emission was subsequently observed in imaging studies of atomic oxygen in flames.[15,16] These observations suggest the occurrence of excitation transfer from the ^3P to the ^5P state (see Fig. 1). The interpretation of these experiments in terms of specific bimolecular rate processes is not straightforward because of the high pressure and chemical complexity of these flames. In addition, at least in the experiment of Miziolek and DeWilde,[1] the tightly focused laser probe was observed to be

promoting multiphoton photolysis of the oxidizer as well as the fuel molecules, leading to a two-photon resonant formation of a microplasma. Nevertheless, even under conditions of no apparent laser probe volume perturbation, it appears that spin-changing collisions of the initially excited level could be a significant collisional removal pathway for laser excited oxygen 3p ^{3}P atoms. In order to employ two-photon excitation for quantitative measurements in combustion environments, such as propellant flames, it is necessary to understand the mechanism for transfer of the initial excitation energy.

This paper reports an experimental study of collisional quenching and excitation trnasfer between the oxygen 3p ^{3}P and ^{5}P states in a controlled low-pressure environment, as well as a companion study of transitions induced between states induced by collisions with a partner in a ^{2}S electronic state. Simple theoretical considerations suggest that open-shell collision partners might be quite effective in inducing spin-changing excitation transfer to the ^{5}P state. Let us take an atom in a ^{2}S electronic state [H(^{2}S), for example] as the simplest open-shell perturber. Approach of a ^{2}S atom to an oxygen atom in the ...$2p^3 3p$ ^{3}P state will yield $^{2,4}\Pi$ and $^{2,4}\Sigma$ molecular curves, while approach to the ...$2p^3 3p$ ^{5}P state leads to $^{4,6}\Pi$ and $^{4,6}\Sigma$ states. We expect the Π molecular curves to have attractive wells and the Σ curves to be mainly repulsive. Furthermore, we might expect a similar qualitative behavior for both the pair of $^4\Pi$ curves and the pair of $^4\Sigma$ curves. Thus, the repulsive $^4\Sigma$ curve correlating with the lower O(3p ^{5}P) + H asymptote will cross the attractive $^4\Pi$ curve from O($3p^3$P) + H. Since the the J·L and L·S terms in the molecular Hamiltonian[17] can mix these two potential energy curves, the $^4\Sigma$-$^4\Pi$ crossing allows a spin-conserving path for collision-induced transitions between the ^{3}P and ^{5}P states, which would not be possible with a closed-shell partner. In our experimental study, we compare the propensity of the stable open-shell oxygen molecule and the closed-shell nitrogen molecule to induce ^{3}P → ^{5}P transitions.

2. EXPERIMENTAL APPARATUS: The apparatus employed for these experiments is shown in Fig. 2. Oxygen atoms were generated by passing a mixture of helium and oxygen (typical partial pressures of 1.2 and 0.2 Torr, respectively) through an Evenson-Broida 2450 MHz resonant cavity. The forward microwave power was kept relatively low (<25 W) in order to minimize the fractional dissociation of the oxygen molecules and production of other transient species. The quenching gases, either N_2 or additional O_2, were added downstream of the microwave discharge. Pressures were measured by a capacitance manometer. The flow tube was pumped by a 50 cfm single-stage mechanical pump.

Two-photon excitation of oxygen atoms was accomplished with tunable uv radiation from a Nd:YAG pumped, frequency doubled and mixed dye laser (Quantel) which operated at a 10 Hz repetition rate. The 226 nm

radiation was separated from other wavelengths by a dichroic mirror; this separation technique was preferred over the use of prisms since, with the former, the laser beam position through the flow system does not change with wavelength. Typical uv power at the apparatus was approximately 1 mJ per pulse, and the bandwidth of the doubled and mixed radiation was ca. 1 cm^{-1} FWHM. The radiation was focused into the center of a flow system with a 30 cm Suprasil lens. No photolytic effects due to laser decomposition of molecular oxygen were evident, as observation of the O atom two-photon signal required the presence of both molecular oxygen and microwave radiation.
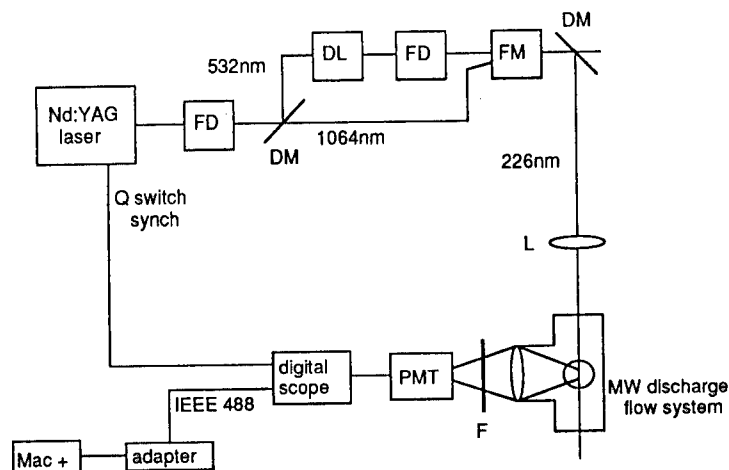

Figure 2.   Schematic diagram of the experimental apparatus.
DL: dye laser; FD: frequency-doubling crystal; FM: frequency-mixing crystal; DM: dichroic mirror; L: lens; F: interference filter;
PMT: photomultiplier tube.

The fluorescence was collected at right angles to the laser beam with a fast (f/2) Suprasil lens and focused through a filter onto a red-sensitive photomultiplier tube (EMI 9516QB). The filters employed were a 850 nm center wavelength, 25 nm bandpass interference filter for detection of the $^3P \rightarrow {}^3S$ emission at 844.7 nm and a 780 nm center wavelength, 10 nm bandpass filter for observation of the $^5P \rightarrow {}^5S$ at 777.5 nm.

The output from the photomultiplier was passed to a boxcar integrator (Stanford Research Systems) for measuring excitation spectra or to a digital oscilloscope (Tektronix 2430A) for capturing the temporal profiles of fluorescence waveforms. The boxcar integrator and oscilloscope were triggered with a synchronization pulse from the Nd:YAG laser Q switch. Some details about the specifications of the

oscilloscope are relevant. The bandwidth of the analog section of this instrument is 125 MHz; the digital sampling rate is 100 megasamples per sec, or 10 nsec between channels. In our experiments we utilized an interpolation feature for repetitive signals, which allowed digitization on a finer grid to achieve the full bandwidth: The waveform from an individual laser pulse was obtained with the usual 10 nsec spacing; however, successive waveforms were taken with a slightly shifted initial delay, controlled internally by the oscilloscope. To improve the signal-to-noise ratio, waveforms were also acquired using a 256-scan running average. Typical emission lifetimes ranged from 40 to 15 nsec, while the laser excitation pulse length was 7 nsec. The waveforms acquired by the digital oscilloscope were transferred through an IEEE 488 interface to a microcomputer (Apple Macintosh Plus) for storage, generation of hard-copy plots, and analysis.

3. EXPERIMENTAL RESULTS: Figure 3 displays typical fluorescence waveforms for emission at 844.7 and 777.5 nm. The noise in these traces is due to the effect of the interpolating feature of the oscilloscope (see Section 2) and the large fluctuations in the fluorescence signal between successive laser shots. It should also be noted that the relative detection sensitivity (filter transmission plus photomultiplier sensitivity, the latter taken from manufacturer's specifications) was larger (4:1) for the collision-induced $^5P \rightarrow {}^5S$ feature.

In addition to the differing intensities for the fluorescent decay of the initially excited ^3P and collisionally populated ^5P states, the two waveforms in Fig. 3 have different temporal profiles. The waveform for the ^3P state in Fig. 3a is the usual exponential curve expected for the decay of an excited level, while that for the ^5P state in Fig. 3b builds up to the peak intensity after an induction period. The time dependence of the ^3P and ^5P populations are expected[2] to obey the following equations:

$$n_3 = n_3(0) \exp(-k_3 t),$$ (1)
$$n_5 = n_3(0) k_{35} [\exp(-k_5 t) - \exp(-k_3 t)]/(k_3 - k_5).$$ (2)

Here, the removal rates k_3 and k_5 include both radiative and collisional processes:

$$k_3 = k_3{}^r + k_3{}^Q(M) [M],$$ (3)
$$k_5 = k_5{}^r + k_5{}^Q(M) [M].$$ (4)

The radiative decay rates are denoted by $k_3{}^r$ and $k_5{}^r$, while the bimolecular quenching rate constants are indicated as $k_3{}^Q(M)$ and $k_5{}^Q(M)$. Equations (3) and (4) are written assuming there is only one quenching species, of concentration [M], present. The quantity k_{35} in Eq. (2) represents that portion of the collisional removal of the ^3P

577

state which results in energy transfer to the ^5P state. We express k_{35} as

$$k_{35} = k_{35}(M) [M],$$ (5)

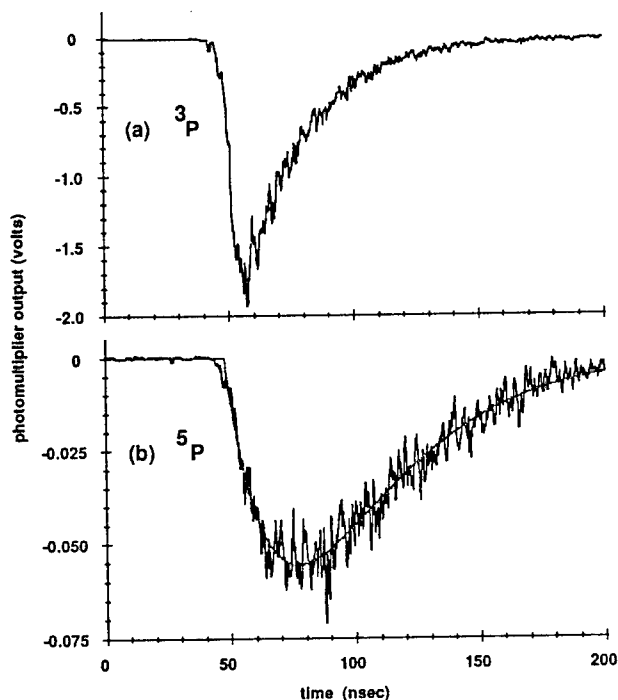where $k_{35}(M)$ is the bimolecular rate constant for $^3P \rightarrow {}^5P$ excitation transfer.

Figure 3. Waveforms for (a) the $^3P \rightarrow {}^3S$ and (b) the $^5P \rightarrow {}^5S$ fluorescence emission upon two-photon excitation of the 3p ^3P state. Pressures: 1.47 Torr He and 0.30 Torr O_2. These waveforms were acquired with the same photomultiplier voltage merely by changing the detection filter.

Equation (1) shows that the ^3P radiative lifetime and quenching rate constants may be determined by measurement of the decay lifetime $\tau = k_3^{-1}$ as a function of quench gas pressure, in the usual Stern-Volmer treatment. The corresponding quantities for the ^5P state must be determined in a nonlinear least squares treatment of the ^5P concentration as a function of time, according to Eq. (2). The time dependence of the ^3P and ^5P concentrations can be determined from the emission intensities at 844.7 and 777.5 nm, respectively.

578

The $^3P \rightarrow {}^5P$ transfer rate k_{35} cannot be determined from analysis of the fluorescence waveforms alone, but rather from the ratio of the integrated emission intensities. The ratio of these integrated emission intensities equals[2]

$$I_5/I_3 = (k_{35}/k_3{}^r)(k_5{}^r/k_5). \qquad (6)$$

The rightmost quantity in parentheses in Eq. (11) is the ^5P fluorescence quantum yield, while the ratio of $k_{35}/k_3{}^r$ represents the ratio of $^3P \rightarrow {}^5P$ energy transfer collisions to ^3P radiative decay events.

Quenching rate constants for the ^3P state were determined from linear least squares fits of the logarithm of the $^3P \rightarrow {}^3S$ emission as a function of time from fluorescence decay curves, similar to that presented in Fig. 3a, at different quencher concentrations. Linear least squares fits, according to Eq. (3), of these data yields the following bimolecular rate constants listed in Table 1 for the total collisional removal rate of the ^3P state. The quoted uncertainties for all rate constants reprted here represent three standard deviations. The collisional and radiative decay rate for the ^5P state was obtained from analysis of $^5P \rightarrow {}^5S$ decay curves such as that presented in Fig. 3b. Because of the weakness of the $^5P \rightarrow {}^5S$ emission signals, this analysis was carried out only for O_2 quencher. The total decay rate k_5 was determined by nonlinear least squares fits[18] using the expected functional form in Eq. (8) for the time dependence of the ^5P population. In this fit, the decay rate k_3 for the ^3P state was fixed at the value calculated from the analysis of the experimental ^3P decay curves. Parameters allowed to vary in this nonlinear fit were the decay rate k_5, an overall normalization constant (proportional to k_{35}), and the channel for which t=0. The solid line in Fig. 3b represents the fit to the experimental data given in that plot. A Stern–Volmer analysis of the derived ^5P rate constants as a function of O_2 density was performed and yields the estimate reported in Table 1 for the bimolecular quenching rate constant by O_2.

Equation (6) shows that measurement of the ratio of the integrated emission intensities for $^5P \rightarrow {}^5S$ and $^3P \rightarrow {}^3S$ fluorescence decay can yield the rate constant k_{35} for collisional transfer from the ^3P to the ^5P state. Accordingly, a series of experiments were carried out to record these waveforms under identical pressure conditions. For a given partial pressure of added quench gas, the fluorescence waveform for one of these radiative decay paths was recorded; then the filter was switched and the other recorded immediately thereafter. The integrated signals were calculated and their ratios corrected for the relative detection sensitivity at 844.7 and 777.5 nm. The integrated intensity ratio was measured as a function of the pressure of both added O_2 and N_2 gas. For both collision partners this ratio is always small and never

increases beyond 5%. The expected dependence of I_5/I_3 on added quencher gas can be calculated from Eqs. (4), (5), and (6). From a fit of I_5/I_3 vs. the added O_2 or N_2 density, the bimolecular excitation transfer rate constants reported in table 1 were obtained.

Table 1. Bimolecular rate constants for quenching and collisional transfer of the 3p ^3P and ^5P states of the oxygen atom at room temperature

Collision Partner	$cm^3/$ (mol-,s)	This Study	Previous Determinations
O_2	k_3^Q	$(7.8\pm0.8) \times 10^{-10}$	$(8.64\pm0.16) \times 10^{-10}$ (Ref. 10) $(6.3\pm0.12) \times 10^{-10}$ (Ref. 8)
	k_5^Q	$(10.8\pm1.8) \times 10^{-10}$	
	k_{35}	6×10^{-11}*	
N_2	k_3^Q	$(5.87\pm0.15) \times 10^{-10}$	$(4.3\pm0.74) \times 10^{-10}$ (Ref. 8) $(2.5\pm0.1) \times 10^{-10}$ (Ref. 4)
	k_{35}	2×10^{-11}*	

*Estimated experimental uncertainty of a factor of 2.

4. THEORETICAL CALCULATIONS: In order to provide a theoretical framework to interpret the experimental excitation transfer rates, quantum scattering calculations[19] have been performed to model oxygen $^3P \rightarrow ^5P$ excitation transfer by the simplest open-shell species, namely the hydrogen atom. Because of the open-shell nature of both collision partners, several potential energy curves arise from each asymptote. Since the high-lying OH potential energy curves are not known, they have been parameterized[3] to be physically reasonable and to be similar to those of known excited alkali hydride curves. With these potentials, the close coupled equations describing the collision process were solved to obtain inelastic cross sections. Initial calculations were carried out for a collision energy $E_{coll} = 208.5$ cm^{-1}, which corresponds to the most probable relative translational energy at room temperature. At this collision energy, the $^3P \rightarrow ^5P$ inelastic cross section was found to be negligibly small. Subsequent calculations were carried out with the $^3P-^5P$ splitting ΔE treated as an adjustable parameter. The calculated cross sections, averaged and summed over the initial and final spin-orbit states, respectively, are reported in Table 2 for various E_{coll} and ΔE. We observe a dramatic dependence on the $^3P-^5P$ splitting, as

580

well as a less pronounced dependence on E_{coll}. While the model potentials may not accurately represent the corresponding OH curves, nevertheless these results suggest that large size of the actual $^3P-^5P$ splitting (2002 cm^{-1}), rather than a lack of coupling between the states causes the excitation transfer rates to be small.

Table 2. Computed degeneracy averaged cross sections for $^3P \rightarrow {}^5P$ transitions induced by hydrogen atoms at various collision energies E_{coll} and assumed $^3P-^5P$ splittings ΔE

E_{coll} (cm^{-1})	ΔE (cm^{-1})	$\sigma_{^3P \rightarrow ^5P}$ (Å^2)
208.5	20	88.7
	60	4.27
	200	3.52
	500	0.40
400	20	142.6
	60	12.6
	200	4.4
600	200	5.18
	500	0.74
2000	60	73.5
	200	12.5
	500	2.60
5000	60	132.55
	200	27.75
	500	6.12

5. DISCUSSION: Table 1 compares our quoted 3p ^3P total quenching rate constants with those determined in previous studies. Our value for O_2 agrees well with that of Bamford, et al.,[10] and is somewhat higher than the most recent result[8] from Kohse-Hoinghaus and coworkers. The present result for the corresponding rate constant for quenching by N_2 is also slightly higher than that of Kohse-Hoinghaus and coworkers.[8] By contrast, the earliest measurement of the N_2 quenching rate constant by Bischel, et al.,[4] is considerably smaller than the present result and that of Ref. 8. We do not know the reason for this discrepancy.

To our knowledge, there have been no previous experimental determinations of collisional properties of the 3p ^{5}P state. While the exact value for the ^{5}P quenching rate constant by O_2 may be uncertain because of difficulties such as those discussed in the previous paragraph, the present experiment nevertheless indicates that this quenching is large and comparable to that for the neighboring ^{3}P state. In attempting to understand the variation of the rate constants for quenching of the nitrogen atom 3p ^{4}D state by the inert gases, Copeland, et al.,[9] noted the importance of the availability of excited states in both the excited atom and the quencher to understand the mechanism of quenching. For the oxygen atom, it is interesting that the overall quenching rate for the 3p ^{3}P and ^{5}P states by O_2 are both large, despite the fact that the former can be removed by excitation transfer to the ^{5}P state, which lies only ca. 2000 cm^{-1} lower, while excitation transfer to lower O atom 3s states for the latter would require removal of ca. 12000 cm^{-1} (see Fig. 1). This suggests that excited states of the O_2 collision partner may be of importance.

The ratio of the rate constants $k_{35}(M)$ to $k_3{}^Q(M)$ equals the fraction f_T of ^{3}P quenching collisions which result in excitation transfer to the lower-lying ^{5}P state. Despite the factor of 2 uncertainty in our measured excitation transfer rate constants, it is nevertheless clear that ^{3}P → ^{5}P collisional transfer explains only a small fraction of the ^{3}P quenching. For O_2 and N_2, we find that f_T equals approximately 8% and 3%, respectively. While f_T is significantly lower for the closed-shell N_2 molecule than for O_2, nevertheless this excitation transfer process is not insignificant for the former. This suggests that the model for this process given in the Introduction may be too simplistic. Moreover, the theoretical calculations with H atoms as the collision partner predict a very small ^{3}P → ^{5}P inelastic cross section.

We surmise that the bulk of ^{3}P quenching collisions occurs by excitation transfer to the collision partner, rather than collision-induced transitions to lower-lying oxygen atom states. This intermolecular excitation transfer would be expected to populate quencher electronic states whose energies lie close to that of the incident ^{3}P state. The initial internal energy of the former is quite high, ca. 88600 cm^{-1} (see Fig. 1), and both O_2 and N_2 possess a number of potential acceptor excited electronic states in this energy range.[20] As the initial excitation energy is greater than the O_2 and N_2 bond energies, dissociative excitation transfer can also occur.

The participation of excited acceptor states in the collisional decay of the ^{3}P state would explain why the model for ^{3}P → ^{5}P collision-induced transitions given in the Introduction is inadequate. Verification of the importance of intermolecular excitation transfer in

582

the quenching of the 3p ^{3}P state could be obtained by observation of excited molecular emission, in the same way that we have shown the existence of intramolecular ^{3}P $\to$ ^{5}P collisional transitions. For example, in the case of N_2 the well-known $C^3\Pi_u$ state has excitation energy very close to that of the O atom ^{3}P state; this electronic state emits in the second positive system in the ultraviolet.[20] Quenching of the ^{3}P state by N_2 can also proceed by a chemical path, namely $O(3p\ ^3P) + N_2 \to NO + N$.

The large uncertainties in the rate constants $k_{35}(M)$ result from the large scatter in the measured integrated intensity ratios I_5/I_3. These fluctuations would be considerably reduced if the integrated signals I_5 and I_3 were measured simultaneously with two photomultiplier detectors, rather than sequentially as in the present experiment. In this way, the intensity ratio would be corrected for shot-to-shot fluctuations in the laser pulse energy and spectral profile. Unfortunately, modification of the apparatus to allow dual photomultiplier detection was not possible in this experimental study.

REFERENCES

1. A.W. Miziolek and M.A. DeWilde, Opt. Lett., Vol. 9, p. 390, 1984.

2. P.J. Dagdigian, B.E. Forch, and A.W. Miziolek, Chem. Phys. Lett., Vol. 148, p. 299, 1988.

3. M.H. Alexander, J. Chem. Phys. (submitted).

4. W.K. Bischel, B.E. Perry, and D.R. Crosley, Chem. Phys. Lett., Vol. 82, p. 85, 1981; Appl. Opt., Vol. 21, p. 1419, 1982.

5. R.P. Lucht, J. Salmon, G.B. King, D.W. Sweeney, and N.M. Laurendeau, Opt. Lett., Vol. 8, p. 365, 1983.

6. M. Alden, H. Edner, P. Grafstrom, and S. Svanberg, Opt. Commun., Vol. 42, p. 244, 1982.

7. M. Alden, A.L. Schawlow, S. Svanberg, W. Wendt, and P.L. Zhang, Opt. Lett., Vol. 9, p. 211, 1984.

8. U. Meier, K. Kohse-Hoinghaus, and Th. Just, Chem. Phys. Lett., Vol. 126, p. 567, 1986.

9. R.A. Copeland, J.B. Jeffries, A.P. Hickman, and D.R. Crosley, J. Chem. Phys., Vol. 86, p. 4876, 1987.

10. D.J. Bamford, L. Jusinski, and W.K. Bischel, Phys. Rev. A, Vol. 34, p. 185, 1986.

11. J.E.M. Goldsmith, Opt. Lett., Vol. 7, p. 437, 1982; J. Chem. Phys., Vol. 78, 1610, 1983.

12. J.E.M. Goldsmith, in: Twentieth Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, p. 1331, 1984.

13. J.E.M. Goldsmith, Appl. Optics, Vol. 26, p. 566, 1987.

14. P.J.H. Tjossem and T.A. Cool, Chem. Phys. Lett., Vol. 100, p. 479, 1983.

15. M. Alden, H. Hertz, S. Svanberg, and S. Wallin, Appl. Opt., Vol. 23, p. 3255, 1984.

16. J.E.M. Goldsmith and R.J.M. Anderson, Appl. Opt., Vol. 24, p. 607, 1985.

17. H. Lefebvre-Brion and R.W. Field, Perturbations in the Spectra of Diatomic Molecules, Academic, New York, 1986.

18. P.R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, New York, 1969.

19. B. Poiully and M.H. Alexander, J. Chem. Phys., Vol. 86, p. 4790, 1987.

20. K.P. Huber and G. Herzberg, Molecular Spectra and Molecular Structure. IV. Constants of Diatomic Molecules, Van Nostrand Reinhold, New York, 1979.

Laser-Based Ignition of H_2/O_2 and D_2/O_2 Premixed Gases Through
Resonant Multiphoton Excitation of H and D Atoms Near 243 nm:
Initial Report of a Deuterium Isotope-Wavelength-Effect in Laser
Ignition

*Brad E. Forch, Dr.
Andrzej W. Miziolek, Dr.
Ballistic Research Laboratory, SLCBR-IB-I
Aberdeen Proving Ground, MD   21005-5066

INTRODUCTION: The use of lasers to initiate combustion events in reactive gaseous mixtures has been the subject of recent investigations in ours, and other laboratories.[1-5] Laser ignition has generally been achieved through photochemical initiation (combustion initiation through chain branching chemical reactions), thermal heating (laser heating of gases), non-resonant spark formation (which results from gas breakdown by intense laser radiation) and through the resonant multiphoton photochemical formation of microplasmas. Laser-induced gas breakdown, which is defined as the point in which single ionization of the gas occurs results from the absorption of many photons which leads to ionization, collision-induced cascade ionization and spark formation. However, laser energy at the threshold of non-resonant gas breakdown is typically far in excess of the requisite minimum ignition energy such that the formation of a blast wave may lead to a detonation.[6]

The resonant multiphoton photochemical formation of microplasmas, which is an ignition source that was developed in our laboratory, appears to be a more controllable means to generate laser-produced sparks than gas breakdown. Recent investigations[1,2] have revealed the first example of a strong wavelength dependence in the amount of incident laser energy which was required to ignite a premixed flow of H_2/O_2 at atmospheric pressure. A tunable ultraviolet (uv) laser system which operates near 225.6 nm was found to induce photodissociation of the oxidizer component, O_2 or N_2O, to produced O atoms in three ground-electronic spin-orbit split states $2p^4\ {}^3P_{2,1,0}$. It was found that the minima in a plot of incident laser energy (ILE) required for ignition versus wavelength were located exactly at the same spectral positions as the O atom two-photon allowed absorption transitions from the $2p^4$ ${}^3P_{2,1,0}$ states to the lowest excited state of the same symmetry.[7-11] Subsequent detailed experimental investigations resulted in the

formulation of a mechanism for this process which consists of three components: (1) the multiphoton photochemical formation of oxygen atoms; (2) resonant multiphoton ionization of these atoms to efficiently form free electrons in the laser focal volume early in the laser pulse; (3) the controlled, resonance enhanced formation of a microplasma using seed electrons which were generated in the previous process.

This ignition method appears to alleviate the problems associated with the sharp and uncontrolled ignition thresholds which are encountered in the extraction of minimum ignition energy measurements using desirable short-pulse lasers (10^{-9} sec) (where energy release occurs within a very short time, over a very small volume in free space and is not associated with catalytic and intrusive effects of electrode surfaces). Therefore, we began an experimental investigation of the potential photochemical interaction of the uv laser and the fuel components of these premixed flows. In this paper, we describe, what is to our knowledge, both the first report of a sensitive wavelength dependence on the laser energy required to ignite a premixed gaseous flow of H_2/O_2 and D_2/O_2 through resonant multiphoton excitation of H and D atoms near 243 nm and the first example of a deuterium isotope-wavelength effect in laser ignition. We show that there is a very definitive excitation wavelength shift near 243 nm (11 cm^{-1}) for the resonant formation of microplasmas that corresponds exactly to H-D deuterium isotope shift of 22 cm^{-1} at the two-photon excitation energy. (D is excited at a wavelength 11 cm^{-1} to the blue of H and the 2S level of D is 22 cm^{-1} higher in energy than the 2S level of H). Plots of ILE versus equivalence ratio, Φ, with the excitation wavelength of the laser tuned to either the H or D-atom two-photon transition, shows a minimum at $\Phi=0.7$ in the fuel lean region at an ILE of ~0.55 mJ. A pressure threshold for microplasma formation was determined and an estimate of the laser power dependence for the ionization process was made which leads to microplasma formation.

EXPERIMENTAL: A schematic of the experimental apparatus that was used in this investigation is given in Fig. 1. Tunable laser radiation near 243 nm was generated by using the second harmonic (532 nm) of a Q-switched Quanta-Ray Nd:YAG laser (DCR-2A,) to pump a Quanta-Ray Dye laser (PDL-1) which was operated at 314.5 nm with a DCM dye. The dye laser beam was then frequency doubled using an angle-cut KDP crystal in the first stage of a servo-motor based tracking system, Quanta-Ray WEX-1; then the doubled radiation was frequency mixed with the 1.06µ fundamental from the Nd:YAG in an angle-cut KDP crystal which was contained in the second-stage, WEX-1 module. The 243 nm laser radiation was separated from unwanted beams with two Pellin-Broca prisms (which were positioned to avoid net beam-steering), a broad-band pass filter and aperture. This configuration yielded 7-nsec (10 Hz), ~2.0 cm^{-1} band width FWHM, 1.5-mJ tunable laser pulses.

The atmospheric pressure burner system consists of a water-cooled and argon-shrouded nozzle with a pinhole aperture (0.6 mm) that was mounted onto an x-y-z translation stage. Gas flows were controlled with Matheson (Model 620) flow meters, which were calibrated by a GCA Precision Scientific wet test meter for H_2, D_2, and O_2 flows up to 2 L/min. Linear flow velocities were in the 10^3-cm/sec range. The incident laser energy was measured just before a 50 mm focal length lens (the beam waist at the focus was about 50μ) with a Scientech (Model 38-0103) disc calorimeter-power-energy meter. The output laser energy could be varied by insertion of one (or more) dielectrically- coated partially transmitting filters in the beam and/or attenuation of the Nd:YAG amplifier stage gain. It was found that the beam characteristics (such as spatial and temporal profiles) were unaltered using this method.

Laser ignition measurements were made by flowing a carefully metered, homogeneous mixture of fuel and oxidizer through the burner. The beam was precisely focused ~0.5 mm above the pinhole orifice, then the laser energy was adjusted until a single pulse ignited the gases and a flame stabilized on the burner surface. The flame was immediately extinguished in order to maintain a constant burner temperature (as not to preheat the gases in succeeding measurements) then the incident laser energy was measured. It was found that the measured values of ILE for ignition were independent of lateral positioning across the burner aperture (entrainment of room air in the Ar-shrouded, premixed gas jet was not occurring). Wavelength dependent ignition plots near 243 nm were generated by holding Φ constant and measuring ignition ILE as a function of excitation wavelength.

Excitation scans of the laser produced microplasmas (in cold gaseous flows or a variable pressure cell) were recorded by collecting emission (with a pair of lenses that were matched to the f/number of a 0.22 m McPherson model 180 monochromator-Hamamatsu R928 photomultiplier system). Emission signals were captured on a 500 MHz Hewlett-Packard model 54111D digital oscilloscope and/or Stanford Research Systems boxcar integrator and strip chart recorder.

Ion signals were detected in a R.M. Jordan time-of-flight mass spectrometer. The system consists of a pulsed-molecular beam valve, a skimmed differenatially pumped laser ionization region, a 1.3 m drift tube and microchannel plate detector. Operation pressure is ~10^{-7} Torr. All gases were Matheson UHP and were used "as is" without further purification.

RESULTS AND DISCUSSION

Microplasma Formation: Figure 2 illustrates the spectroscopic
schemes which have been utilized for the detection of hydrogen atoms in
flames, molecular beams, discharges and other environments.[12-17] The
energy gap between the 1S ground electronic state and the 2S lowest
excited electronic state (near-degenerate $2S_{1/2}$, $2P_{1/2}$, $2P_{3/2}$ levels) is
~10 eV, Fig. 2, scheme 1. Since tunable laser sources which operate in
this regime are not readily available, two-photon absorption of laser
radiation at 243.07 is required to excite the 1S-2S transition.
Absorption of an additional third photon from the resonantly excited 2S
state is sufficient to ionize an H atom since the three-photon
excitation energy (3 x 5.1 eV) exceeds the 13.6 eV H atom ionization
potential. (Note that absorption of an additional photon by an H atom
in any one of the excited state electronic manifolds given in Fig. 2,
schemes 1-5, are sufficient to achieve ionization.) Detection of laser
induced fluorescence from resonance excitation of the n=2 level in
flames and other environments is difficult since the emission at 122 nm
is absorbed by flame and room gases. Therefore, we investigated the
possibility of inducing multiphoton photolysis of H_2 in a cold
atmospheric flow, followed by resonance enhanced (2+1) multiphoton
ionization H atoms to form a laser-produced microplasma. Within the
high temperature plasma environment, there exist collisional processes
which may populate (for example) the n=3 or 4 levels where Balmer-β
(486.1 nm) or Balmer-α (656.3 nm) emissions could be detected. Figure
3a gives an excitation wavelength scan where Balmer-β emission was
monitored. The signal maximum is found precisely at the known two-
photon resonance transition of atomic hydrogen, although the spectral
width is much larger than that of a purely atomic transition. This
observation is clearly indicative of a much more complex, highly non-
linear phenomena. i.e., the formation of a microplasma. The width of
the excitation spectrum is entirely consistent the finite absorption
cross section in the wings of the transition. Multiphoton absorption
and ionization in the spectral wings of the atomic transition generates
free electrons (priming or seed electrons) early in the laser pulse
which initiate cascade ionization and the formation of a plasma which is
heated up a very high temperature by the inverse brehmsstrahlung effect.

The temporal profile of the emission, Fig. 4b, is also indicative
of microplasma formation. The lifetime of the emission (~40 ns) is much
greater than that of the time-resolved scattered laser radiation
(~7 nsec), Fig. 4a, or, for example, from resonance multiphoton laser
induced fluorescence from H-atoms that we have detected using scheme 5,
where the emission signal follows the temporal profile of the exciting
laser pulse, Fig. 4a. Although microplasma formation is initiated by
atomic ionization, the lifetime of the emission is purely a property of
the microplasma.

Figure 5a–d, gives the emission profiles as a function of four different laser pulse energies (0.15–1.5 mJ). Note that as the excitation energy increases, the widths of the spectral profiles likewise increases and shows a linear dependence on laser pulse energy, Fig. 6. Extrapolation of the measured linewidths to zero laser energy give a FWHM of ~6 cm^{-1}, which is a factor of about three greater than the measured spectral bandwidth of the laser at this wavelength. It is known, for example, that at pressures near atmospheric the radiation intensities required to cause breakdown by the cascade processes are sufficiently strong that optically induced Stark shifts (which is a linear function of laser energy) and broadening of the electronic levels of an atom occur and have a marked influence on transition probabilities.[16,18] The signals certainly are severely AC Stark broadened, however the widths are much greater than the expected natural linewidth of the atomic transition ($\ll 1$ cm^{-1}). This evidence further indicates that emission we observe is not simply an atomic property, but that of a much more complex event.

To further verify the importance of resonance enhancement in multiphoton photochemical formation of a microplasma, these experiments were repeated under identical conditions except that H_2 was replaced with D_2. Figure 3b gives the normalized spectral profile of emission signal which was observed. Note that spectra a and b are essentially identical except that the emission signal in b reaches a maximum at a wavelength position which is ~11 cm^{-1} to the blue of a. (The spacing at the two–photon energy is ~22 cm^{-1}). The observation wavelength–bandpass of the monochromator was broadened and centered to equally catch both emissions at 656.3 nm and 656.1 nm for H and D respectively. A simple calculation and check with a known value[19] shows that this separation is nearly identical with the H–D isotope shift (the D 2S level is 22.4 cm^{-1} higher in energy than the H 2S level). These results strongly support our conclusion that microplasma formation is the result of inital photolysis of H_2 or D_2 to produce ground electronic state H and D atoms and subsequent multiphoton excitation and ionization are important in the efficient wavelength dependent formation of these microplasmas, just as had been observed previously for oxidizers, O_2 and N_2O.[1,2]

Laser Power Dependencies: In order to gain information on the laser power dependence for the photolysis and MPI process we performed a series of experiments in a well–controlled low–pressure (10^{-7} Torr) environment using a molecular–beam time–of–flight mass spectrometer, Fig. 1. We were unable to detect any H^+ or D^+ from laser irradiation of molecular beams of H_2 or D_2 when the laser beam was focused into the ionization region of the spectrometer (with the tightest focusing (200 mm f.l. lens) that geometrical constraints would allow). We did, however, find a threshold of ~70 Torr, ILE=1.2 mJ, for the onset of microplasma formation in a variable pressure flow cell, marked by a

arrow in Fig. 7. These results suggest that a collisionally induced photodissociation process may be responsible for the production of ground state atoms and microplasma, and is currently under investigation. We were able to generate intense pulsed and effusive H and D atom beams with a hot W filament inserted before the skimmed differentially pumped ionization region in the TOF-MS.[20] Excitation spectra, Fig. 8, were recorded by mass gating either the H^+ or D^+ signals with a boxcar integrator, then scanning the laser through the two-photon resonance, three photon ionization transitions (2+1) (note that there was virtually no signal off resonance of either excitation wavelength). Again, a wavelength spacing of 11 cm^{-1} is evident, which corresponds to a deuterium isotope shift of 22 cm^{-1} at the two-photon energy). Laser power dependencies for ion signal production were made, Fig. 9, assuming a signal dependence on laser intensity of I^n, where n equals the number of photons absorbed. Values of n = 2.75 ± 0.24 and 2.63 ± 0.36 were obtained for H^+ and D^+ production, respectively, which corresponds to two-photon resonance excitation and one photon ionization at low laser energies (<0.5 mJ). These results (as expected) indicate that the overall excitation and ionization was a three-photon process. The non-integer values are indicative of partial saturation of the initial two-photon absorption and ionization step. We could easily fully saturate the ionization process at higher laser energies and record laser energy dependencies of ~2 for H^+ and D^+. Although the dissociation energy of H_2 is ~4.5 eV and the energy of a photon at 243 nm is ~5.2 eV, absorption of a single photon is insufficient for the photoproduction of an H atom from H_2 because of negligible single-photon absorption at this wavelength (the ionization potentials of H_2 and O_2 are 12.06 and 15.43, respectively). The overall laser energy dependence for the microplasma formation must then be at least a five photon process and clearly is highly nonlinear. In comparison, photoproduction of atomic oxygen from molecular oxygen photodissociation near 225.6 nm gave a two-photon dependence even though the single photon energy would be sufficient for photolysis.

Ignition Experiments: We then began a series of experiments on premixed flows of H_2/O_2 and D_2/O_2 at atmospheric pressure in order to investigate the possibility of using these microplasmas as ignition sources. Figures 10a-b show the wavelength dependence on the amount of ILE necessary to ignite premixed flows of H_2/O_2 and D_2/O_2 ($\Phi=0.7$). The curves clearly show a strong dependence of the ILE on the laser wavelength. Two prominent minima are evident and correspond exactly to the spectral locations of the H and D atom two-photon resonance excitation wavelengths near 243 nm as described above. Apparently, the focused uv laser (the spot size of the laser beam was about 1.96 x 10^{-5} cm^2 so that the laser power was about 4 x 10^9 W cm^{-2}) near 243 nm not only photodissociates H_2 or D_2 to yield ground state atoms, but also, when on resonance with H or D, requires the least amount of laser

energy to ignite the gases into a stabilized combustion. We believe that this is the first report of the ignition of these reactive gases from the resonant formation of a microplasma through multiphoton excitation of H and D atoms. Furthermore, we believe that this is the first report of a deuterium isotope-wavelength effect in laser ignition. The wavelength shift of +22 cm^{-1} of the D_2/O_2 ignition curve relative to that of the H_2/O_2 curve is clearly related to the H-D deuterium isotope shift and underscores the importance of resonance enhancement in the ignition event.

A comprehensive characterization of the ignition phenomena reveals that the ion formation is a key feature in resonance multiphoton microplasma ignition since it is a means by which electrons can be formed in a very controlled and efficient manner early in the laser pulse. These "seed" electrons are responsible for the growth of the microplasma (spark) which if intense enough and heated to a sufficiently high temperature permit transition into full combustion. It is important to reiterate that this technique offers the ability to control the amount of laser energy required for ignition and it may potentially circumvent the threshold problems associated with non-resonant laser induced breakdown to initiate combustion. A gas mixture which is initially transparent to the laser beam requires substantially much higher energy fluxes to generate seed electrons which serve as the basis of the formation of a plasma. However this much higher energy flux may greatly exceed the ignition threshold so that a very intense plasma and blastwave may result, which may lead to a detonation. We found that we could form microplasmas very easily through resonance ionization, at very low laser energy, that were not intense enough to cause ignition. Thus, our method allows for precise control of laser energy in the ignition process. By comparison, we have investigated ignition of these gas mixtures using a 532 nm laser beam (which is initially transparent to the mixture) under identical experimental conditions and found that: (1) a factor of nearly forty times more laser energy was required for ignition; (2) the 532 nm beam dependence of ILE on equivalence ratio is flat (independent of Φ; (3) a relatively intense spark was produced. This behavior is entirely consistent with our mechanism and reasoning.

Figure 11a-b gives plots of incident laser energy required for ignition of premixed flows of H_2/O_2 and D_2/O_2 versus Φ when the laser is tuned to the two-photon resonance excitation wavelengths for H and D atoms respectively. The observed minima in both curves occurs in the fuel lean region ($\Phi=0.7$, ILE=0.55 mJ). Apparently this behavior, which was also observed in previous work,[1-3,21] is characteristic of light, diffusive fuels such as H_2, which can replenish burned fuel early on in the evolution of the ignition kernel as it expands into a stabilized flame front. In our recent work on excimer laser (ArF, 193 nm) ignition

593

of H_2/O_2 we also found that the most efficient ignition was in fuel lean region.[22]

CONCLUSION: We have observed a strong wavelength dependence on the ignition of H_2/O_2 and D_2/O_2 premixed flows using a tunable laser near 243 nm. Furthermore, we have observed a deuterium isotope shift in both microplasma formation in H_2 and D_2 gases and in the ignition of H_2/O_2 and D_2/O_2 flows. These results results underscore the importance of two-photon atomic resonances in the microplasma formation process. We have demonstrated that our recent observations of a new laser ignition phenomenon that involves resonant multiphoton photochemical formation of microplasmas appears to be more general and applies to fuel molecules (H_2 and D_2) as well as oxidizers (O_2 and N_2O). We believe that this is the first report of a sensitive wavelength dependence on the laser energy required to ignite these mixtures through resonant multiphoton excitation of H and D atoms (produced from H_2 and D_2 photolysis) and the first report of a deuterium isotope effect in laser ignition. Measurement of the ILE required for ignition versus equivalence ratio (Φ) shows that the most efficient ignition occurred with 0.55 mJ ILE at $\Phi=0.7$ in the fuel lean region. Strong experimental evidence is given which shows that ignition occurs through the resonant formation of a laser-produced microplasma in a well-defined volume. These new experimental results indicate that resonance enhancement in the formation of a microplasma is a well controlled ignition method which appears to alleviate the problems associated with the sharp thresholds encountered in the well-known laser-produced spark (gas breakdown) process. Currently we are exploring other possible resonance effects for the purpose of activating or enhancing the combustion of other reactive systems.

REFERENCES

1. B.E. Forch and A.W. Miziolek, Opt. Lett., Vol. 11, p. 129, 1986.

2. B.E. Forch and A.W. Miziolek, Comb. Sci. and Tech., Vol. 52, p. 151, 1987.

3. J.A. Syage, E.W. Fournier, R. Rianda, and R.B. Cohen, J. App. Phys., Vol. 64, p. 1499, 1988.

4. M. Lavid and J.G. Stevens, Comb. and Flame, Vol. 60, p. 195, 1985.

5. B. Raffel, B. Warntaz, and J. Wolfrum, Appl. Phys., Vol. B37, p. 344, 1985.

6. F.J. Weinberg and J.R. Wilson, Proc. R. Soc. London Ser. A, Vol. 321, p. 41, 1971.

7. M. Alden, H. Edner, H.P. Grafstrom, and S. Svanberg, Opt. Comm., Vol. 42, p. 244, 1982.

8. A.W. Miziolek and M.A. DeWilde, Opt. Lett., Vol. 9, p. 390, 1984.

9. U. Meier, K. Kohse-Hoinghaus, and Th. Just, Chem. Phys. Lett., Vol. 126, p. 567, 1986.

10. J.E.M. Goldsmith, J. Chem. Phys., Vol. 78, p. 1610, 1983.

11. P.J. Dagdigian, B.E. Forch, and A.W. Miziolek, Chem. Phys. Lett., Vol. 148, p. 299, 1988.

12. R.P. Lucht, J.T. Salomon, G.B. King, D.W. Sweeney, and N.M. Laurendeau, Opt. Lett., Vol. 8, p. 365, 1983.

13. M. Alden, A.L. Schawlow, S. Svanberg, W. Wende, and P.L. Zhang, Opt. Lett., Vol. 9, p. 211, 1984.

14. J.E.M. Goldsmith, Opt. Lett., Vol. 7, p. 437, 1982.

15. J.E.M. Goldsmith, Twentieth Symposium (International) on Combustion, p. 1331, The Combustion Institute, 1984.

16. P.J.H. Tjossem and T.A. Cool, Chem. Phys. Lett., Vol. 100, p. 479, 1983.

17. B.E. Forch, J.B. Morris, and A.W. Miziolek, "Laser-Based Approaches in Luminescence Spectroscopy", invited book chapter, Laser-Induced Fluorescence and Ionization Techniques in Combustion Diagnostics, (T. Vo Dinh and D. Eastwood, Ed.), in press, 1990.

18. P. Lambropoulos, Phys. Rev. A., Vol. 9, p. 1992, 1974.

19. A.R. Striganov and N.S. Sventitskii, Tables of Spectral Lines of Neutral and Ionized Atoms, p. 75, IFI/Plenum, 1978.

20. S.W. Downey and R.S. Hozack, Opt. Lett., Vol. 14, p. 15, 1989.

21. B. Lewis and G. von Elbe, Combustion, Flames and Explosions of Gases, p. 390, Academic Press, Inc., 1951.

22. B.E. Forch, A.W. Miziolek, C.N. Merrow, J.B. Morris and R.J. Locke, Proceedings of the 1989 JANNAF Combustion Meeting, The Chemical Propulsion Information Agency, in press.
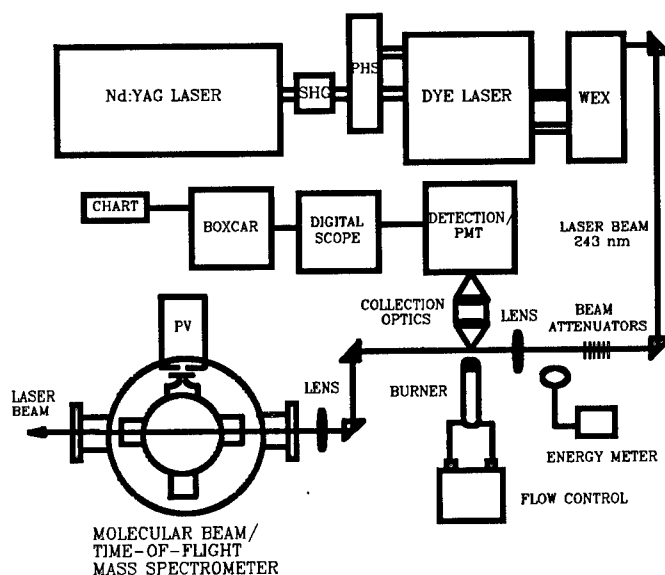
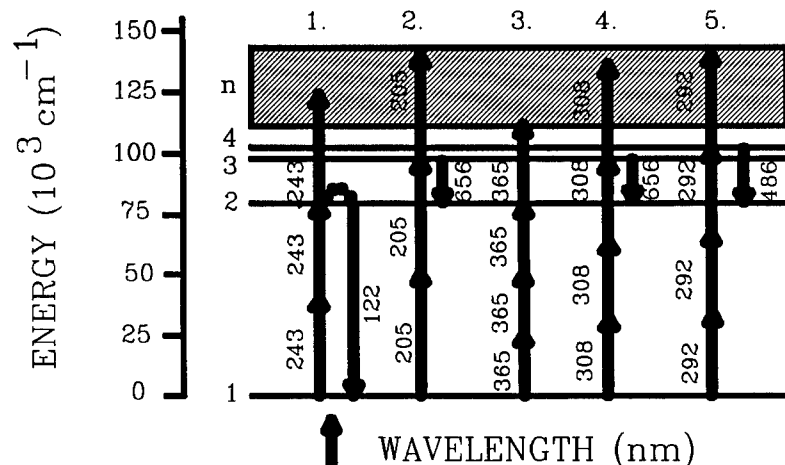Fig. 1. Schematic of experimental apparatus.

Fig. 2. Simple energy level diagram for atomic hydrogen. Schemes 1-5 depict various laser-based multiphoton processes for excitation and ionization of the H atom.
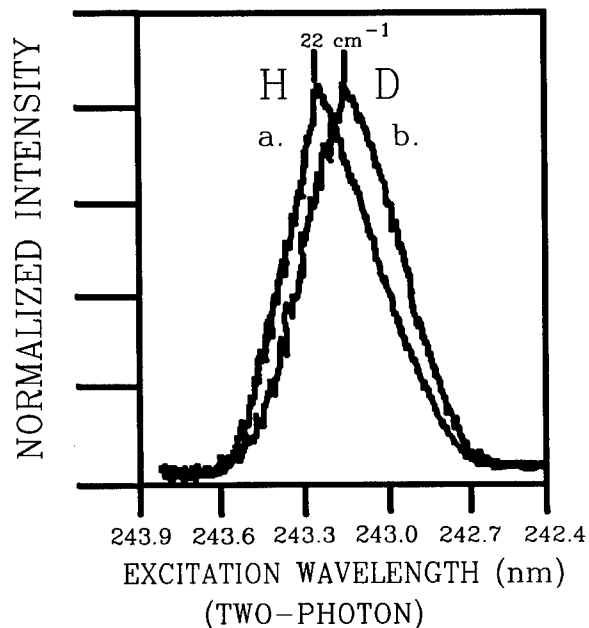

Fig. 3. Excitation wavelength scans for microplasma formation near 243 nm with the laser beam focused (ILE ~1.2 mJ) into cold flows of: (a) H_2, λ obs.=656.3 nm and (b) D_2, λ obs.=656.1 nm. The wavelength separation between peaks is 11 cm^{-1} which corresponds to 22 cm^{-1} a the two-photon excitation energy which is identical with the H-D deuterium isotope shift.
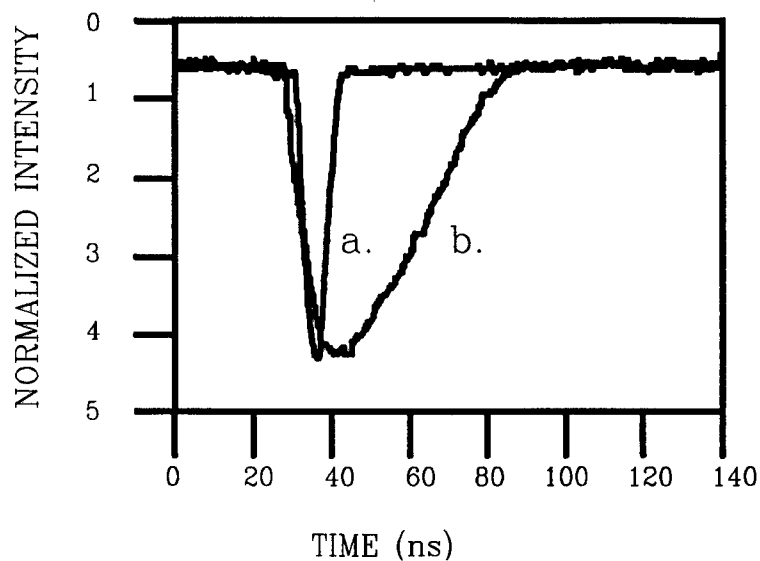
597

Fig. 4. Time resolved emission from: (a) scattered laser radiation and (b) H-atom emission at 656.3 nm from microplasmas formed in H_2 flows, using scheme 1. Fig 2, ILE = 1.2 mJ.
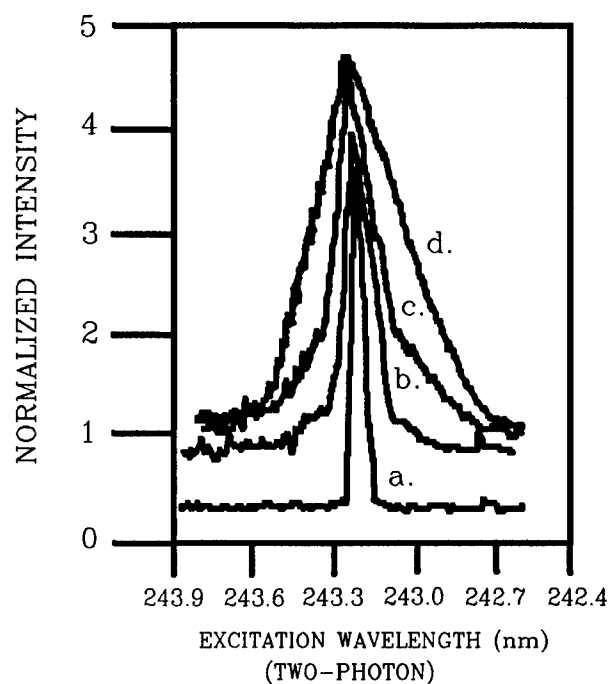
Fig. 5. Excitation wavelength scans for microplasma formation near 243 nm plotted as a plotted as a function of ILE: (a) 0.15 mJ, (b) 0.43 mJ, (c) 0.73 mJ and (d) 1.4 mJ in at 300K in an atmospheric pressure flow.
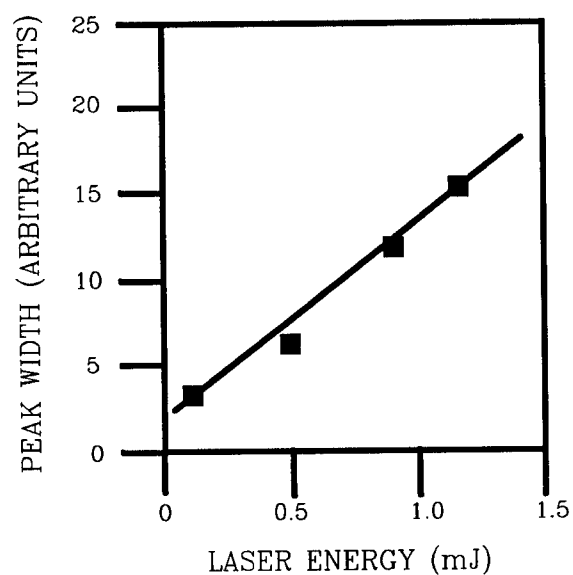
598

Fig. 6. Plot of the spectral width of the emission signals in Fig. 5 versus ILE.
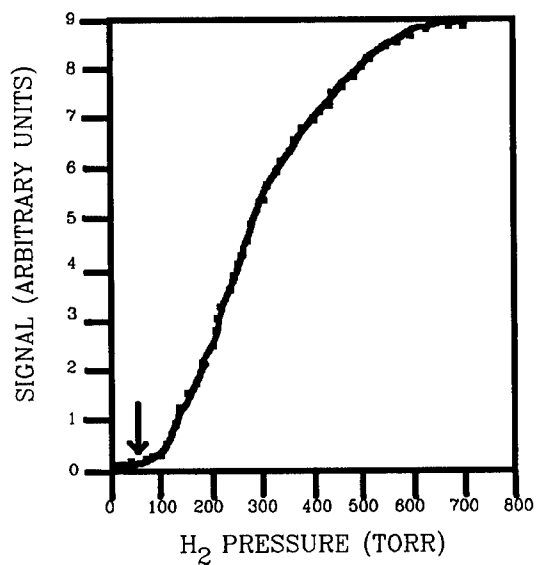
Fig. 7. Plot of H-atom emission signal at 656.3 nm from a microplasma versus H_2 pressure in a variable pressure cell, ILE=1.2 mJ.

599
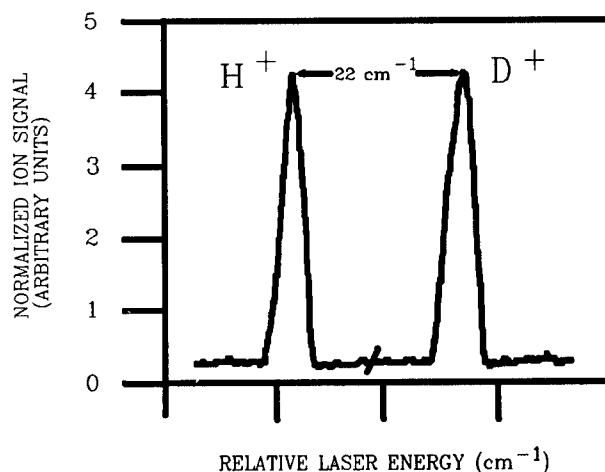
Fig. 8. Excitation spectra of H^+ and D^+ taken in a time-of-flight mass spectrometer using resonance two-photon excitation and three-photon ionization (2+1) of H and D atoms near 243 nm. The wavelength separation between peaks is 11 cm^{-1} which corresponds to 22 cm^{-1} a the two-photon excitation energy which is identical with the H-D deuterium isotope shift.
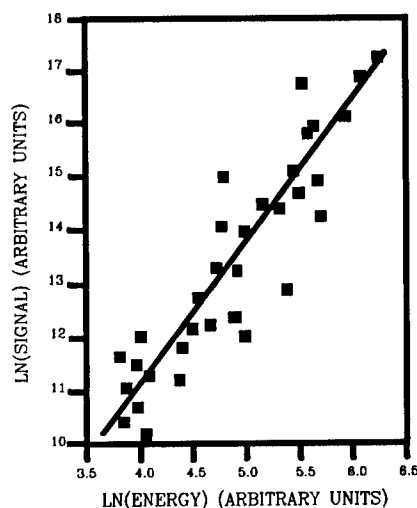
Fig. 9. Laser energy dependence plot for H^+ production from a beam of H atoms in a time-of-flight mass spectrometer. A value of n=2.75 ± 0.24 was obtained from the slope in a linear least squares fit (solid line).
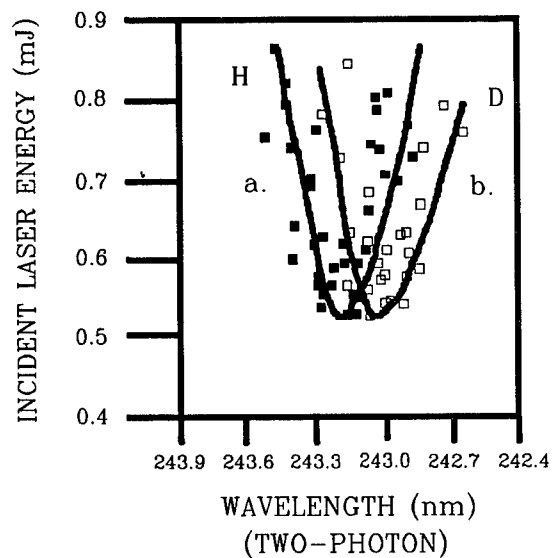
Fig. 10. ILE necessary to ignite premixed flows of: (a) H_2/O_2 and (b) D_2/O_2, as a function of excitation wavelength near 243 nm. A shift of +11 cm^{-1} of ignition curve (b) relative to ignition curve (a) is evident.
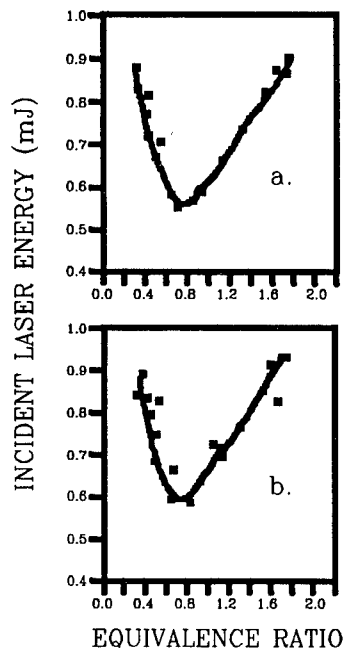
Fig. 11. Dependence of the ILE required to ignite premixed flows of: (a) H_2/O_2 and (b) D_2/O_2, as a function of equivalence ratio. The laser was set at the peak of the two-photon excitation wavelength of: (a) H atoms at 243.07 and (b) D atoms at 243.00.

Laser-Based Multiphoton Excitation Processes in Combustion
Diagnostics

*Brad E. Forch, Dr.
Andrzej W. Miziolek, Dr.
Jeffrey B. Morris, Dr.
Ballistic Research Laboratory, SLCBR-IB-I
Aberdeen Proving Ground, MD 21005-5066

INTRODUCTION: Light atomic species such as oxygen and hydrogen
atoms are fundamentally important in a wide variety of combustion
processes such as flame ignition, propagation and extinction and in
chemical flame reactions. The hydrogen atom/radical, for example, is
necessarily important in nearly every combustion process because of its
high reactivity and diffusivity, yet its direct detection therein is
very difficult. The oxygen atom is of equal importance in pertinent
flame processes, flame structure and relevant chemical and mass transfer
processes but its direct detection is equally difficult. The nitrogen
atom is a constituent of many gaseous, liquid and solid oxidizers (of
particular importance to the chemistry of propellant combustion) since
it is chemically bound in nitrate, nitro and nitroso compounds, but has
never been detected in a flame environment (to our best knowledge).
Conventional optical detection of these light atoms is difficult for a
variety of reasons. In particular, the necessary optical resonance
excitation wavelengths fall far into the ultraviolet (uv) and the
requisite tunable laser sources are generally not available for such
flame atomic spectroscopy. Historically, several different laser
spectroscopic techniques have been employed for both detection and
diagnostics of these light atoms in combustion processes.[1-7] One of the
most successful methods thus far for flame diagnostics has been laser-
induced fluorescence (LIF).[3] Since the energy gap between the ground
and excited electronic states of the light atoms is typically on the
order of 10 or more electron volts (ev), techniques such as two-photon
or multiphoton laser excitation methods must be employed. Multiphoton
excitation of these species can be conveniently provided with
commercially available laser sources that can quite easily generate the
"one-step" photons that may be used in the "multi-step" excitation
process.

The advent of tunable, narrow-linewidth lasers has also enabled the development of another multiphoton technique for the detection of and quantification of species such as light atoms, namely multiphoton ionization (MPI).[8,9] Similar to multiphoton laser induced fluorescence experiments, ionization of the atom can be achieved if the energy of the total number of photons which are absorbed is sufficient to eject an electron from the target species, i.e., greater than the ionization potential (IP). Consequently, the "free electrons" which are liberated or positive ions which are produced are amenable for detection. Flame atomic species have been detected, as such, with resonant multiphoton optogalvanic spectroscopy.[4-7] Optogalvanic signals are detected in our laboratory using a platinum tipped probe (described in the experimental section of this paper) which is a point source detector and well suited for spatial measurements in combustion environments. A particular advantage of optogalvanic detection over LIF is that dark or non-emitting species may be studied. However, both the MPI and MPE approaches suffer from limitations. MPE inherently suffers from quenching considerations and signal collection inefficiencies. A recent paper by Meier, et al.,[10] addresses this issue by measuring the efficiency of fluorescence quenching by the various species present in the detection region. Aside from quenching considerations, the technique of MPI, in comparison, lacks selectivity in signal detection and thus is quite susceptible to background ionization.[3] There also exists the possibility of introducing pertubations as a result of inserting the probe into the flame medium. Furthermore, multiphoton excitation techniques for flame light atoms may suffer from complications which arise from laser photolysis due to the uv wavelengths, tight beam focusing and high peak laser powers which are required and utilized to drive the highly non-linear processes.[11]

Two other diagnostic techniques have been used for direct detection of atomic species, namely Raman scattering[12] and coherent anti-Stokes Raman spectroscopy.[13] Both have been employed for flame detection and diagnostics, but suffer from sensitivity limits and other spectral interferences that have prevented further development of these spectroscopies for atomic flame detection.

We are particularly interested in the application of lasers for detailed flame chemistry studies and have been in the process of building up a comprehensive and well-equiped research program in this area. This paper presents an extension on our previous work in light atom diagnostics. We have applied the techniques of MPE and MPI to a variety of flames, in particular H_2/O_2, H_2/air, H_2/N_2O, C_2H_2/O_2 and C_2H_2/air, for the purpose of further development and implementation of diagnostic techniques for small atomic species such as H, O, C, and N in combustion environments. A great deal of information on, and insight into the details of the various excitation processes can be gained by

604

measuring laser power dependencies. These power dependencies were measured for the photoexcitation of O atoms in H_2/O_2 and H_2/N_2O flames as well as for the photolytic production and photoexcitation of O atoms in cold flows of the oxidizer components of these flames. Similar investigations were performed on H atoms. Laser power dependencies were measured for the photoexcitation of hydrogen atoms in atmospheric pressure flames and in a molecular-beam time-of-flight mass spectrometer. In the course of our experimental work we have also become very sensitive to the potentially intrusive nature of the focused laser as a flame diagnostic probe due to photochemical effects caused by the intense uv pulses. We will discuss the effect of photochemical pertubations introduced by lasers in light of our combustion diagnostic work on oxygen atom; in particular, as related to H_2/O_2 premixed flames and reactive gaseous flows. We have investigated the potential utility of the photolytic nature of the focused uv laser with applications to ignition and combustion enhancement by resonant multiphoton photochemical means.[14,15]

EXPERIMENTAL SECTION: The experimental apparatus used in the flame diagnostic work will be described. Tunable laser radiation was provided by a 10 Hertz Quanta-Ray Q-switched Nd:YAG-pumped dye laser system (DCR-2A, PDL-1) which is equipped with a wavelength extension system (WEX-1). The WEX-1 is a servo-motor based tracking system which houses two modules which contain non-linear crystals (angle-cut KDP) that are used to frequency double the dye laser output and frequency mix it with the fundamental (1064 nm) of the Nd:YAG to generate tunable wavelengths in the deep ultraviolet (uv). The unwanted beams were separated from the desired uv beam with a Pellin-Broca prism, broad-band filters and an aperture. Typical outputs from the laser system are: two milliJoules (mJ) of uv energy at 226 nm using a doubled rhodamine 590 dye and mixed (two-photon excitation of O atoms, $2p^3\ 3p^3P \leftarrow 2p^4\ ^3P$), one mJ at 243 nm, using a doubled DCM dye and mixed (two photon excitation of H atoms, n=1→2), 10 mJ at 292 nm, using a double Kiton Red dye (three-photon excitation of H atoms n=1→4), 15 mJ at 308 nm, using an acid shifed (HCl) Rhodamine 640 dye (three-photon excitation of H atoms n=1→3), 5 mJ at 365 nm, using a base shifted (saturated NaOH) Fluorescene 548 and mixed (three photon excitation of H atoms n=1→2) and 70 µJ at 193.1 nm, doubled Kiton-Red, mixed and second anti-Stokes H_2 Raman-shifted (one photon excitation of carbon atoms $^1D_2 \rightarrow\ ^1P^0_1$). The measured laser linewidth in the uv was 2.2 cm^{-1} with a temporal width of about 5 nsec. The pulse energy was measured with a Scientech volume-absorbing disc calorimeter (Model 38-0103) and analog meter.

A simple atmospheric pressure jet burner system was fabricated from a Swagelock 0.635-cm stainless steel terminator fitting through which a 0.5 mm hole was drilled. The fuel and oxidizer were premixed in the burner and the gas flow was regulated by Matheson (Model 620) flowmeters

605

and metering valves that were calibrated with a GCA Precision Scientific wet test meter. For H_2 and O_2, flows up to 2 liters/min were used which resulted in linear flow velocities in the 10^3-cm/sec range. The burner was equipped with a water cooling jacket and shroud of inert argon gas which encircled the flame and minimized mixing with air and recirculating flow. The laser induced fluorescence was collected at right angles to the exciting light with a pair of fused silica biconvex lenses and imaged onto the adjustable slits of a 0.22 m McPherson monochromator (Model 218) which were shielded with the appropriate broad-bandpass and/or interference filters. The photomultiplier (Hamamatsu R928 or EMI 9659QA) signal was then fed to either a Princeton Applied Research or Stanford Research Systems (SRS) boxcar integrator then plotted on a strip-chart recorder, or fed to either a (model 7912AD) Tektronix transient digitizer (500 MHz) or Hewlett Packard (Model 54111D) digital oscilloscope (500 MHz) in order to capture the temporal profile of the emission. Excimer laser radiation (at 193 nm) was provided by either a Lumonics Hyper-EX Model 440 or Model TE 860 both of which were operated in the unstable resonator mode.

Ions of both atomic and molecular species were detected at atmospheric pressure using an optogalvanic probe which could be forward biased to detect electrons liberated in the laser induced multiphoton photoionization process or reversed biased to detect the resultant positive ionic species. The probe was designed, constructed, characterized and supplied by Professor Terrill A. Cool of Cornell University.[7]

Mass spectra were recorded with an R.M. Jordan time-of-flight mass spectrometer. The system is equipped with a pulsed-molecular beam valve which could be operated near 10^{-8} torr.

RESULTS AND DISCUSSION:

O-ATOMS:

MPE. Two-photon absorption from a uv laser beam to excite ground state oxygen atoms to the first excited state of identical spin symmetry was first reported by Bischel, et al.[17] They described a scheme which is illustrated in Figure 1. The photoexcited ^3P state ($2p^3 3p\ ^3P \rightarrow 2p^4\ ^3P_2$) can radiate in the near infrared (at 844.7 nm) to the ^3S state, which is a useful analytical wavelength for the detection of the emission. Oxygen atoms were first detected in a realistic combustion environment by Alden, et al., utilizing this scheme.[2]

We have also utilized this excitation scheme for oxygen detection in our experimental work.[11,14-16] We have recorded the fluorescence excitation scan of ground state O atoms in the primary reaction zone of

a premixed stoichiometric (Φ=1.0) H_2/O_2 atmospheric pressure flame. The ground spin-orbit split states, J=0, 1, 2 are clearly resolved although the excited state splittings were not resolved with the resolution of our laser system. The signal to noise ratio was excellent (100:1). The signal strength was found to be highly sensitive to the flame reaction zone that was probed. The signals were strong, although heavily quenched, and followed the temporal profile of the exciting laser beam. We measured the laser power dependence for the process (by either varying the gain of the amplifier stage of the Nd:YAG laser or inserting quartz plates into the 226 nm beam to attenuate the energy) and found that there was a quadratic dependence, i.e., n = 1.9 $\pm$ 0.2 (the uncertainty was obtained from the standard deviation of the mean from three measurements). Experiments were also performed in a premixed H_2/N_2O and H_2/air flames and identical experimental results were obtained. As previously reported, there was a strong collisional energy transfer to the ^5P state of oxygen with detectable fluorescence emission at 777.5 nm which was also utilized to perform similar diagnostic measurements.[11]
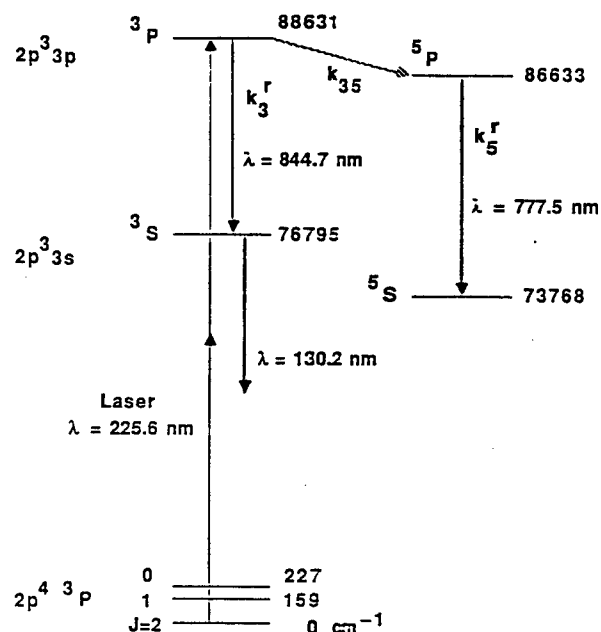
Figure 1. Simple Energy Level Diagram for Oxygen Atoms. Terms are defined in the text.

MPI. Atomic oxygen was also detected in these flame environments using an optogalvanic probe. Here the probe was inserted into the flame itself and the laser was focused within <0.5 mm of the platinum anode of the detector (the burner head itself was grounded). An excitation scan was recorded by scanning the laser through the three ground spin-orbit split states while using the probe in the forward biased mode where free

607

electrons which were liberated in the photoionization process could be collected. The laser power dependence for the process was measured as described earlier. At low laser energy (0.1 to 0.5 mJ) we observed a slope of 2.8 ± 0.3 for n which is indicative of an overall three-photon photoionization process. At higher laser energy (0.6 to 1.2 mJ) a value for n = 2.1 ± 0.2 was obtained which probably indicates saturation of the two-photon step in the 2+1 resonance ionization process. A similar saturation effect was observed by us in our hydrogen atom work as well as recently by Downey and Hozack[18] who observed n change from 3.2 to 2 in a 2+1 photoionization process at power densities of near 10^7 W/cm^2.

Photolytic Effects. Earlier experimental results from our laboratory first indicated that in diagnostic work on oxygen atom detection in a flame environment there was evidence that the focused uv laser and high peak powers that were used led to laser induced multiphoton photolysis of the parent oxidizer molecule N_2O, yielding unusually high concentrations of oxygen atoms (higher than that which would normally be found in a combustion environment).[11] The manifestation of this phenomenon, which also occurs in O_2-containing systems, is clearly illustrated in Figure 2 which gives a normalized intensity profile of atomic oxygen atoms in a premixed H_2/O_2 flame ($\Phi=0.8$). In Figure 2a, the laser was tuned to and fixed at the J=2 ground spin-orbit split state; the observation wavelength was 777.5 nm, a 300 mm focal length lens was used and the laser energy was 0.35 mJ. It was found that the signal intensity was very sensitive to the positioning of the laser focal spot within the flame environment and reached a maximum at a distance of 3.5 mm above the burner head (the starred point in Figure 2), which was the location of the primary reaction zone, then rapidly fell away as the probe region was translated into the post-flame gases. In Figure 2b, the same experiment was then repeated, however this time a shorter focal length lens was used (100 mm) and the laser energy was increased to 0.8 mJ. Very large signals were observed throughout the flame structure. However, as the the sampled flame region was translated into the post-flame gases it was found that the O atom signal intensity did not drop off rapidly as it had in the previous experiment, Figure 2a. Apparently, the focused uv laser beam was inducing the photodissociation of molecular oxygen to produce excess quantities of atomic oxygen. Furthermore, in a darkened laboratory when a tighter focusing lens was used (50 mm) it was quite evident that the formation of a tiny microplasma was being induced, and in addition a faint acoustic report was audible. Even when the laser energy was attenuated to the point wherein microplasma formation was not evident, it was found that relatively high O atom signal intensities could be found. It is known that molecular oxygen does not absorb substantially at wavelengths > 200 nm at room temperature.[19] However, at these flame temperatures (post-flame gas temperature of ca. 1500 K) O_2 ground state vibrational levels are thermally populated such that

single photon excitation at 226 nm through these states (Schumann-Runge bands) which are highly predissociative can lead to the production of excess quantities of O atoms which subsequently may be two-photon excited and emit. Goldsmith recently described a two-laser experiment where he scanned across molecular-oxygen Schumann-Runge bands at 221 nm in a flame environment and monitored O atom production with a second laser tuned to the J=2 state at 225.6 nm and found a similar photolytic production of excess quantities of atomic oxygen.[20]
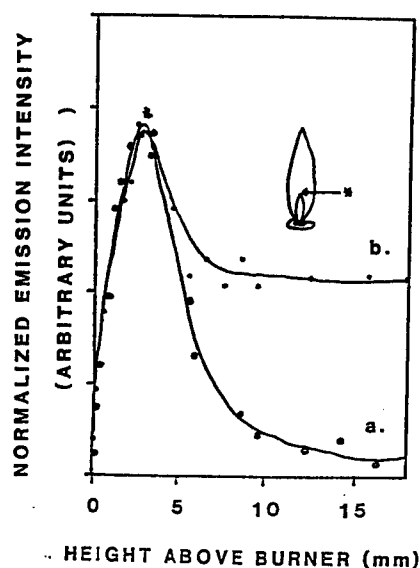
Figure 2. (a) Oxygen Profile Atom in an H_2/O_2 Atmospheric Pressure Flame and (b) Same Profile Taken at Higher Laser Energy and Shorter Focal Length Lens as Described in the Text. The starred point (*) indicates the location of the primary reaction zone.

We have noted in our flame diagnostic work, not only the photolytic production of O atoms in flames, but also in cold gas flows of O_2, N_2O or air when no flame was present at all. Apparently, the focused laser was inducing the photolysis of these gases to produce ground state oxygen atoms which could then be easily two-photon excited and emission could be detected at 777.5 or 844.7 nm. We decided to determine the laser power dependence n for the process by measuring the signal intensity as a function of laser energy in the cold flows. We obtained a value of n = 3.4 ± 0.3 for an O_2 flow and a similar value for an N_2O flow. This result suggests that a two-photon process is responsible for

the photolytic production of atomic oxygen from either O_2 or N_2O in a room temperature flow. The actual value obtained is n = 3.4 - 2.0 = 1.4 (the non-integer value is most likely due to some saturation of the two-photon excitation photolysis processes and/or losses due to rapid photoionization which competes with the fluorescence). Excitation spectra of O atoms were taken with the optogalvanic probe (free electrons or positive ions were collected in each scan respectively). At low laser energies we measured a value of n = 4.3 ± 0.3 for the overall process: photolysis of O_2 to produce O atoms (two photons), photoexcitation of O atoms (two photons) and the photoionization of O atoms (1.0 photons) to yield detectable electrons. The overall process probably requires 5 photons (not 4.3).

Microplasma Formation. The experimental results in the above section clearly indicate that laser photolysis effects can potentially lead to erroneous measurements of species profiles in combustion environments. We have demonstrated that there are indeed wavelength specific interactions with parent fuel and oxidizer photofragments thatlead to resonant ionization, the liberation of free electrons within the laser focal volume and ultimately the formation of a laser produced microplasma.[14,15] We have investigated the feasibility of the utilization of these laser produced microplasmas as a technique for the initiation of combustion events and have studied in some detail the chemistry and mechanism involved in this process. Specifically, we have used the uv laser as an ignition source for premixed H_2/O_2 and H_2/N_2O flows at atmospheric pressure. This was the first report of such a strong and sensitive wavelength dependence for the initiation of combustion in a reactive gaseous flow. A sharp wavelength dependence on the amount of incident laser energy required to ignite these mixtures was reported. The wavelength dependence exhibits a spectral profile similar to the two-photon excitation curve for flame oxygen atoms. Figure 3 gives the amount of incident laser energy (ILE, the energy measured before the focusing lens) required to initiate the combustion event as a function of laser wavelength. Ignition occurred within the shaded areas. The respective peaks correspond exactly to the same spectral positions as the oxygen atom two-photon wavelengths near 225.6 nm. Clearly, oxygen atom production and subsequent excitation is an important step in the efficient (~0.5 mJ ILE required) laser ignition of H_2/O_2 flows in this wavelength region. Time resolved measurements were performed on O_2 and N_2O flows alone and it was found that the resonant formation of a microplasma (with a lifetime on the order of 100 nsec) was responsible. A new combustion initiation phenomenon was discovered which involves atomic resonances in the uv which we have named "multiphoton photochemical microplasma ignition." The mechanism for microplasma formation consists of three major components: (1) the multiphoton photochemical formation of oxygen atoms; (2) multiphoton ionization of these atoms to efficiently form free electrons in the

laser focal volume and; (3) the formation of a laser microplasma using the electrons formed in the previous process. The major difference between this and previous work on laser spark formation and gas ignition is that the multiphoton photochemical resonance ionization process provides a more efficient and controllable means of liberating free electrons in the laser focal volume which leads to the laser spark formation.
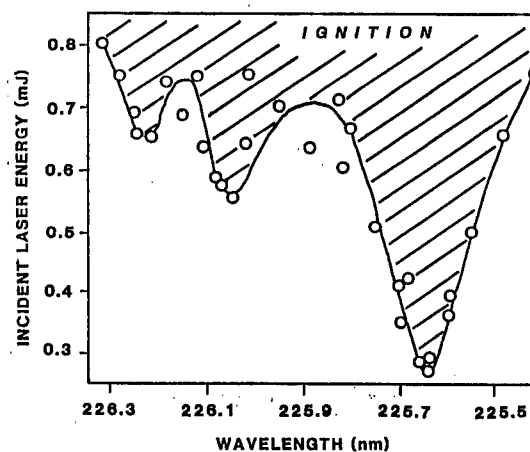

Figure 3. Plot of the Incident Laser Energy Required to Ignite a Premixed Flow of H_2/O_2 as a Function of Laser Wavelength in the Region of 226 nm

H-ATOMS:

MPE. We have also been interested in the detection of hydrogen atoms (the smallest of atoms) because of its extreme importance in nearly every combustion environment and relevant chemical reactions. Many multiphoton excitation schemes have been implemented by various investigators to detect H atoms in a wide variety of flames,[1,3,5,7,10,21] see Figure 4. The first reported measurement of H atom concentrations in a flame by two-photon excited fluorescence was reported by Lucht, et al.,[1] in a premixed $H_2/O_2/Ar$ low pressure flame (20 Torr), scheme 2, Figure 4.

We attempted to use the three-photon excitation, scheme 4, Figure 4. Absorption of three photons at 308 nm from the n=1 to the n=3 state is followed by observation of emission at 656.3 nm. This particular wavelength is particularly attractive as a diagnostic technique since it coincidentally falls within the spectral gain of a XeCl excimer laser,

611

wherein hundreds of mJ of laser energy are potentially available. However, we were not able to detect H atom emission since there was strong molecular interference from the hydroxyl radical (OH).
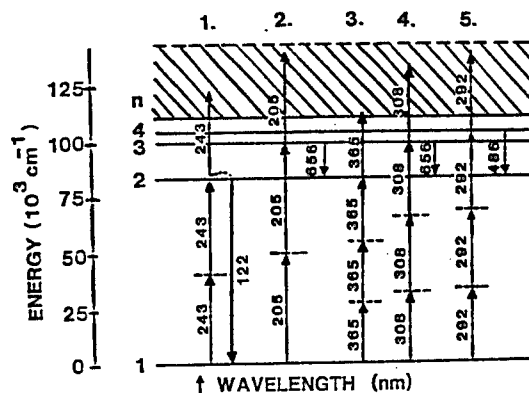

Figure 4. Simple Energy Level Diagram and Multiphoton Excitation Schemes for H Atoms. Energy on the vertical scale is given in wavenumbers. The arrows correspond to photon wavelengths.

H atoms were successfully detected in an atmospheric pressure H_2/O_2 flame using scheme 5, Figure 4. Ground state H atoms were excited from the n=1 to the n=4 level through a three-photon absorption at 292 nm. Fluorescence emission at 486 nm due to the n=4 to the n=2 radiative process (Balmer-β emission) was detected. Figure 5a gives an excitation scan recorded as the uv laser was scanned through this three-photon resonance in a stoichiometric H_2/O_2 atmospheric pressure flame while monitoring the emission with a monochromator/filter combination. Figure 5b gives the fluorescence emission scan. We measured the laser power dependence for the three-photon excitation process. Five independent measurements were made and we obtained a value of n = 2.77 ± 0.4 for H atoms in the primary reaction zone. This measurement is consistent with the expected value of n=3. This excitation scheme may prove to be an excellent choice for diagnostic measurements of H atom profiles in combustion environments since strong fluorescence signals could be obtained quite easily (even at atmospheric pressure) and because of the convenient diagnostic wavelength available for the detection of the emission.
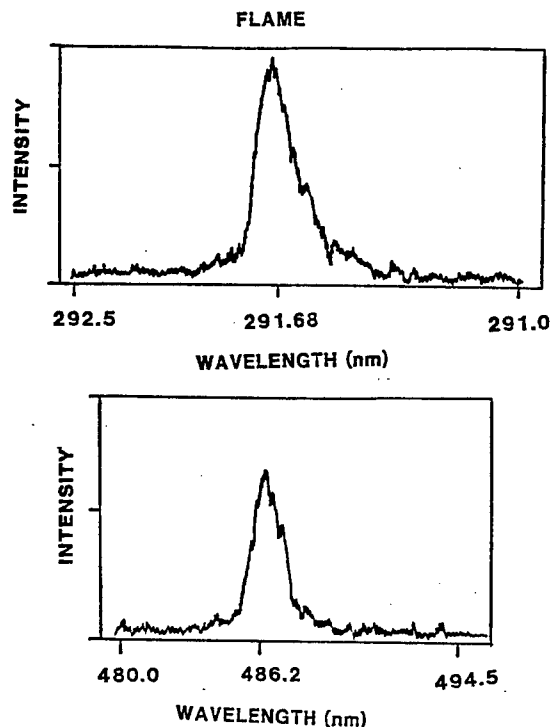
FLAME

Figure 5. (a) Excitation Spectrum for H Atoms Using the Three-Photon Excitation (λ=292 nm) Scheme 5 as Depicted in (b). The emission spectrum taken at 486 nm from the three-photon excitation process (λ=292 nm) for H atoms.

MPI. H atoms could also be detected using excitation scheme 3, Figure 4, in cold gaseous flows of H_2 or D_2 and in an atmospheric pressure H_2/O_2 fuel rich flame. H atoms in the flame environment were excited through two-photon excitation from the ground state n=1 to the lowest excited state n=2. We could induce multiphoton photolysis of the parent fuel in the preheat or post-flame gases in a combustion environment or cold flow of H_2, ionize the atoms and form a laser produced microplasma. Within the high temperature plasma environment there exist collisional excitation and de-excitation processes which produce atomic hydrogen in the n=3 excited state such that the radiative decay at 656.3 nm can be detected. This experiment was repeated with a flow of D_2 and an isotopic blue shift of 22 cm^{-1} was observed in the excitation spectrum which is in excellent agreement with the well known value of 22.4 cm^{-1}. In order to gain information on the laser power dependence for the MPI process, we generated an H or D molecular beam with a hot wire filament in a time-of-flight mass spectrometer.

Excitation spectra were recorded by mass gating either the H^+ or D^+ signal with a boxcar integrator and then scanning the uv laser beam through the two-photon transition. Very strong ionization signals (several volts) were seen using the dual microchannel plate detector (mcp) which was located at the top of a 1.5 m drift tube. A value of n = 2.75 ± 0.2 and 2.63 ± 0.3 was obtained for H and D, respectively, corresponding to two-photon excitation and single photon ionization at low laser energies (<0.5 mJ). These results clearly indicate that the overall excitation and ionization process was a three-photon process.

N- and C-Atoms: Here we will briefly discuss our first attempt at the detection of atomic nitrogen and carbon atoms in a combustion environment using tunable uv light. To the best of our knowledge, neither nitrogen atoms nor carbon atoms have ever been detected in a combustion environment using any optical diagnostic technique. Nitrogen atoms were first detected utilizing a two-photon excitation scheme by Bischel, et al.[17] They produced ground-state N atoms in a low pressure microwave discharge then excited the atoms through a two-photon transition at 211 nm ($2p^3 {}^4S^0 \rightarrow 2p^2 3p\ {}^4D^0$). Fluorescence emission in the near infrared (869 nm) from the radiative process ($2p^2 3p\ {}^4D^0 \rightarrow 2p^2 3s\ {}^4P$) was detected. We tried to utilize this excitation technique to detect ground-state nitrogen atoms in atmospheric pressure fuel lean H_2/N_2O and CH_4/N_2O flames but the signal was swamped by background interference which remains unidentified. We also attempted to detect carbon atoms in an atmospheric pressure stoichiometric flame by generating tunable radiation in the wavelength region of 193 nm to excite carbon atoms from the metastable ${}^1D_2$ state to the ${}^1P^0_1$ at 193.1 nm. This metastable state may be formed through some elementary reaction rate process or by laser photolysis.[22] A suitable wavelength for the detection of the emission is the ${}^1P^0_1 \rightarrow {}^1S_0$ radiative process at 247.9 nm. We obtained ~70 μJ of tunable laser radiation but found that there were "holes" in the output of the uv beam as it was scanned throughout this wavelength region. It appears that absorption of molecular oxygen through the Schumann-Runge $B^3\Sigma^-_u \leftarrow X^3\Sigma^-_g$ band system bleached holes in the output profile of the laser beam. These are the dominant absorption bands for O_2 and, unfortunately, the R(17) and P(15) absorption bands overlap exactly with the wavelength which corresponds to the carbon atom resonance at 193.1 nm.[19]

CONCLUSION: We have investigated the potential utility of the application of multiphoton excitation processes for the detection of light atomic species such as O, H, N, and C atoms in combustion environments. Techniques which were utilized include multiphoton laser-induced fluorescence and ionization, the latter using an optogalvanic probe. Laser power dependencies were determined for most of the excitation schemes which were implemented. These measurements gave detailed information on the mechanism in the photoexcitation process

such as the number of photons involved, potential photolytic effects which occured therein and information on saturation effects. It was determined which of these schemes were most successful for the detection of each specific atom in a combustion environment. Clearly multiphoton excitation is a viable approach for the detection of these light atomic species in combustion environments.

We discussed how the requisite high peak-powers that were utilized in the multiphoton excitation schemes could photochemically produce excess concentrations of atomic species (higher than that normally found in a combustion environment) and lead to erroneous concentration profile measurements. It was described how these photochemical pertubations could be used to augment combustion events such as ignition. This led to the discovery of a new ignition phenomenon which has been named "multiphoton photochemical microplasma ignition." The detailed steps involved in laser produced microplasmas have been determined and further laser ignition studies are currently under investigation.

REFERENCES

1. R.P. Lucht, J.T. Salmon, G.B. King, D.W. Sweeney, and N.M. Laurendeau, "Two-Photon-Excited Fluorescence Measurement of Hydrogen Atoms in Flames," Optics Letters, Vol. 8, No. 7, pp. 365-367, July 1983.

2. M. Alden, H. Edner, P. Grafstrom, and S. Svanberg, "Two-Photon Excitation of Atomic Oxygen in a Flame," Optics Communications, Vol. 42, No. 4, pp. 244-246, July 1982.

3. M. Alden, A.L. Schawlow, S. Svanberg, W. Wendt, and P.L. Zhang, "Three-Photon-Excited Fluorescence Detection of Atomic Hydrogen in an Atmospheric-Pressure Flame," Optics Letters, Vol. 9, No. 6, pp. 211-213, June 1984.

4. J.E.M. Goldsmith, "Resonant Multiphoton Optogalvanic Detection of Atomic Oxygen in Flames," Journal of Chemical Physics, Vol. 78, No. 3, pp. 1610-1611, February 1983.

5. J.E.M. Goldsmith, "Resonant Multiphoton Optogalvanic Detection of Atomic Hydrogen in Flames," Optics Letters, Vol. 7, No. 6, pp. 437-438, July 1982.

6. J.E.M. Goldsmith, "Flame Studies of Atomic Hydrogen and Oxygen Using Resonant Multiphoton Optogalvanic Spectroscopy," Twentieth Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, pp.1331-1335, 1984.

7. P.J.H. Tjossem and T.A. Cool, "Detection of Atomic Hydrogen in Flames by Resonance Optogalvanic Spectroscopy," Chemical Physics Letters, Vol. 100, No. 6, pp. 479-482, September 1983.

8. P. Lambropoulos, "Theory of Multiphoton Ionization: Near-Resonance Effects in Two-Photon Ionization," Physical Review A, Vol. 9, No. 5, pp. 1992-2013, May 1974.

9. G. Mainfray and C. Manus, "Resonance Effects in Multiphoton Ionization of Atoms," Applied Optics, Vol. 19, No. 23, pp. 3934-3940, December 1980.

10. U. Meier, K. Kohse-Hoinghaus, and Th. Just, "H and O Atom Detection for Combustion Applications: Study of Quenching and Laser Photolysis Effects," Chemical Physics Letters, Vol. 126, No. 6, pp. 567-573, May 1986.

11. A.W. Miziolek and M.A. DeWilde, "Multiphoton Photochemical and Collisional Effects During Oxygen-Atom Flame Detection," Optics Letters, Vol. 9, No. 9, pp. 390-392, September 1984.

12. C.J. Dasch and J.H. Bechtel, "Spontaneous Raman Scattering by Ground-State Oxygen Atoms," Optics Letters, Vol. 6, No. 1, pp. 36-38, January 1981.

13. R.E. Teets and J.H. Bechtel, "Coherent Anti-Stokes Raman Spectra of Oxygen Atoms in Flames," Optics Letters, Vol. 6, No. 10, pp. 458-460, October 1981.

14. B.E. Forch and A.W. Miziolek, "Oxygen-Atom Two Photon Resonance Effects in Multiphoton Photochemical Ignition of Premixed H_2/O_2 Flows," Optics Letters, Vol. 11, No. 3, pp. 129-131, March 1986.

15. B.E. Forch and A.W. Miziolek, "Ultraviolet Laser Ignition of Premixed Gases by Efficient and Resonant Multiphoton Formation of Microplasmas", Combustion Science and Technology, Vol. 52, pp. 151-159, February 1987.

16. P.J. Dagdigian, B.E. Forch, and A.W. Miziolek, "Collisional Transfer Between and Quenching of the 3p ^{3}P and ^{5}P States of the Oxygen Atom," Chemical Physics Letters, Vol. 148, No. 4, pp. 299-308, July 1988.

17. W.K. Bischel, B.E. Perry, and D.R. Crosley, "Detection of Fluorescence from O and N Atoms Induced by Two-Photon Absorption," Applied Optics, Vol. 21, No. 8, pp. 1419-1429, April 1982.

18. S.W. Downey and R.S. Hozack, "Saturation of Three-Photon Ionization of Atomic Hydrogen and Deuterium at 243 nm," Optics Letters, Vol. 14, No. 1, pp. 15-17, January 1989.

19. M.P. Lee and R.K. Hanson, "Calculations of O_2 Absorption and Fluorescence at Elevated Temperatures for a Broadband Argon-Fluoride Laser Source at 193 nm," Journal of Quantitative Spectroscopy and Radiative Transfer, Vol. 36, No. 5, pp. 425-440, May 1986.

20. J.E.M. Goldsmith, "Photochemical Effects in Two-Photon Excited Fluorescence Detection of Atomic Oxygen Atoms in Flames", Applied Optics, Vol. 26, No. 17, pp. 3566-3572, September 1987.

21. T.W. Hansch, S.A. Lee, R. Wallenstein, and C. Wieman, "Doppler-Free Two-Photon Spectroscopy of Hydrogen 1S-2S*," Physical Review Letters, Vol. 34, No. 6, pp.307-309, February 1975.

22. R.C. Sausa, A.J. Alfano, and A.W. Miziolek, "Efficient ArF Laser Production and Detection of Carbon Atoms from Simple Hydrocarbons," Applied Optics, Vol. 26, No. 17, pp. 3588-3593, September 1987.

A Very High Shock, Self-Contained Data
Acquisition System

*Raphael A. Franco, Jr., Dr.
James K. Ingram, Mr.
US Army Engineer Waterways Experiment Station
Vicksburg, Mississippi 39180-6199

BACKGROUND

Acquisition of dynamic test data began in earnest in the 1930's with mechanical scratch-plate recorders which evolved into electromechanical pen-type recorders. A natural evolution resulted in higher frequency devices which depended on oscilloscopes with either drum-type strip film recorders or photosensitive paper recorders during the 1940-60's. The third generation of recorders welcomed a large advance in capabilities with high-fidelity magnetic tape recorders which had their heyday in the 1970's. Emerging digital technology brought on the present, fourth generation, systems which began to make serious inroads in the early 1980's. Not to be overlooked is an equivalent advancement in transducer design, signal conditioning/amplifier electronics, and data processing methodologies. The advantages of digital systems are many, and include: fast software command and control, high frequency response, high resolution, and direct compatibility with digital signal processing algorithms.

Although lower performance, shock-hardened, self-recorders have been developed[1,2], to our knowledge, none have resolution above 8-bits, frequency response above 10-kHz, or on-board memory exceeding 32-kwords. None of the other devices are designed for connecting several units together with a common master clock for multi-channel operation when no time skew is allowed. The competing recorders also do not possess simple sensor shunt calibration capability.

REQUIREMENT

The emerging and rapidly advancing field of precision penetrator weapons research requires the capability to accurately measure the extreme inertial forces and stresses present in the test environment and on the

test article subjected to this environment. An equally severe requirement is that of measuring explosion generated shock effects very near to the source of a detonation to support advanced weapons effects research. Survival of the data acquisition system required to measure environment/ performance has been the most elusive criterion to meet. Although transducer technology has for several years allowed the measurement of accelerations, stresses, and pressures to extremely high levels, the normally required signal cable transmission link is highly vulnerable, and is totally impractical for use with projectiles fired from high velocity guns or rocket assist. Impact decelerations and explosion-produced accelerations typically exceed 100,000-g's in current scenarios.

<div align="center">CONCEPT</div>

The concept for a workable, stand-alone data acquisition system concentrated on integrating the sensor, signal conditioner, amplifier, and data recorder into a single, miniature package (Figures 1, 2, and 3).
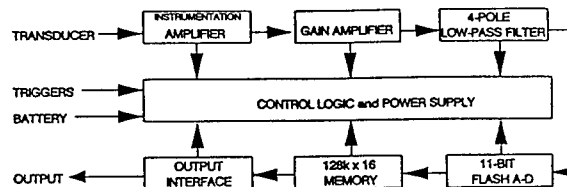
Figure 1. HDAS functional block diagram.

Benefits provided by the new system are:

1. Measurement of phenomena at levels not otherwise possible.
2. Elimination of signal cables.
3. Elimination of cable-generated noise/cross-talk.
4. Greatly enhanced frequency response.
5. Report quality scaled engineering data plots and integrated plots immediately after recovery of the modules.

Minimum requirements for such a system included:

1. Very small size/mass.
2. Very low power consumption.

Side 1. Signal conditioner/amplifier.

Side 2. Control logic/digital memory.

Figure 2. HDAS circuit board components.

Figure 3. Unitized canister/sensor/recorder system for projectile impact tests.

3. High shock-hardness.
4. Specifications/performance equivalent to current high-resolution magnetic tape systems and state-of-the-art transient data recorders.

The smallest practical size/mass is required to maintain the necessary high frequency response and to minimize the damaging effects of the near-instantaneous loading produced 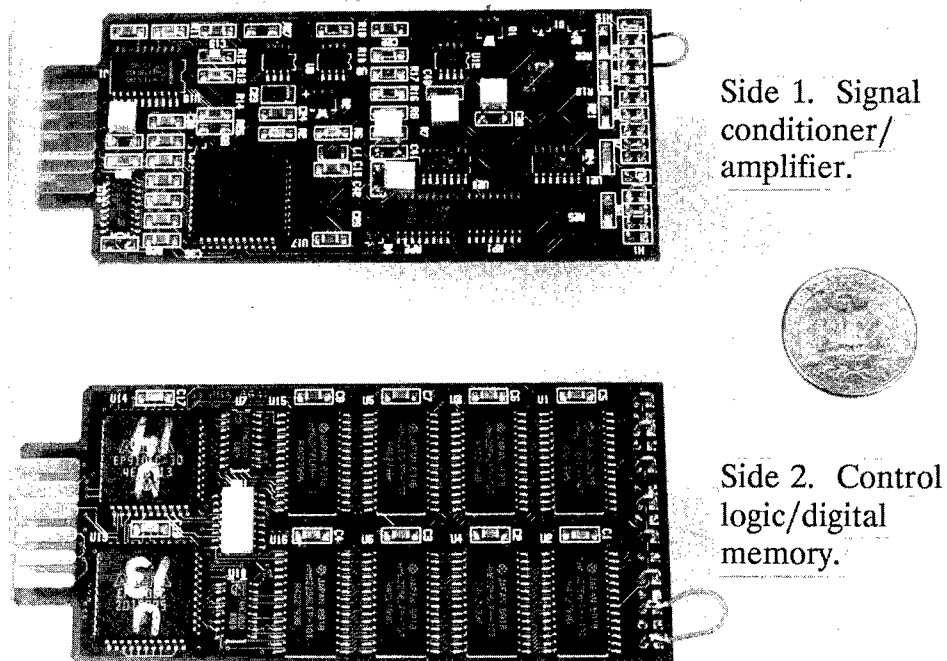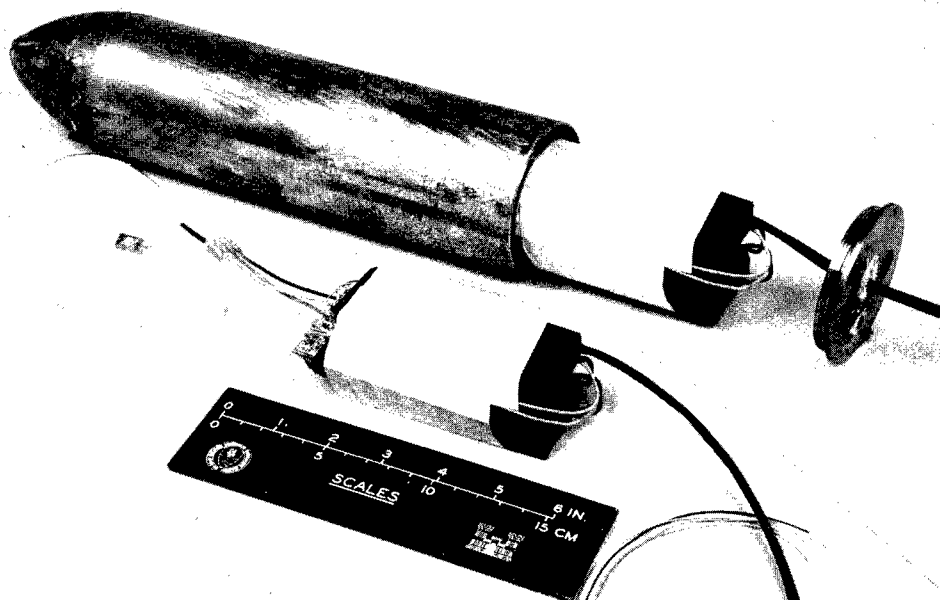by the severe inertial fields that are generated. The mechanical frequency response of the measurement system is inversely proportional to the size of the system, i.e., the smaller the size/mass, the higher the frequency response. Typical acceleration loads are on the order of 100,000- to 200,000-g's. Future research requirements will likely extend the survival level criterion to even higher magnitudes.

Initial requirements were for single point measurement of projectile performance. Other requirements necessitate measurements at multiple points within the penetrator weapon in order to more precisely characterize its response. This requirement mandates use of several channels of data acquisition, further compressing the overall size constraint. Laboratory penetrator projectiles are miniature devices that pose an additional, severe size limitation on autonomous data acquisition systems.

This limits not only the size of the electronics, but also the size of the battery power supply. Size limitations imposed on the battery limit the amount of available power, which mandates a highly-efficient, low power drain system. This requirement is further enforced by the requirement for the system to maintain memory for as long as several months after data acquisition, while recovery of the system is in progress. For penetrator weapon tests, such as tactical aircraft runway penetrators, multiple tests will be conducted over a period of months before the runway is excavated for recovery of the recorder systems. For large weapons effects tests, the recorder systems must be placed in the test bed as much as 2- to 3-months prior to test execution, and occasionally must remain in place from one to several weeks before recovery to allow for crater and test article digout, and surface ejecta analysis.

Since both the electronics and battery power supply are now integrated with the sensor inside the test article, all components are exposed to the same severe shock loading environment. Bare circuit boards, even with hardened components, invariably fail in the test environments being evaluated. All system components must survive the test environment, not only during data acquisition, but through recovery and data read-out. Additional shock hardening must be added to ensure both physical and electronic function survivability.

System performance specifications require a minimum flat frequency response at least equal to that of the currently available transducers (100-kHz). This specification dictates a sampling frequency of at least 1-MHz (1-usec/sample) for a 10-point sine wave resolution. In order to maintain adequate dynamic range, a minimum 11-bits is required (this is the current limit of monolithic analog-to-digital converters (ADC's) having required speed, shock hardness, and low power specifications). Sufficient on-board memory is required to ensure full capture of the phenomena of interest (at least 100 msec at the maximum sampling rate, 1-MHz).

Table 1 lists the electrical and physical characteristics of the current, validated HDAS system.

TABLE I. HDAS Technical Specifications

ELECTRICAL CHARACTERISTICS

Number of Channels 1

Input .... Instrumentation Amplifier

Filter .... 4-Pole Butterworth

Bandwidth .... DC to 160kHz

Sampling Rate (Selectable)............ 100kHz to 1MHz

Amplitude Resolution .... 2048 Digital Steps

Storage Capacity .... 128k Words (128 x 16)

Pre-Trigger Storage .... 12k Words

Calibration Data .... 4k Words

Triggers .... Switch Closure or 5 Volt Pulse

Supply Voltage .... 8 to 15 Volts

Supply Current

    Standby .... 100 Microamps (Typical)
    Data Acquisition .... 135 Microamps (Typical)
    Data Retention .... 100 Microamps (Typical)

PHYSICAL CHARACTERISTICS

Temperature Range .... 0 to +70 Celsius

Maximum Shock Level .... ± 150,000g's

Recommended Sensor .... Piezoresistive Bridge

Recorder

    Length 110mm (4.33in)
    Width 54.94mm (2.15in)
    Height 13.97mm (0.56in)
    Weight 71gm (2.49oz)

Battery Pack

    Length 99.06mm (3.90in)
    Width 35.56mm (1.40in)
    Height 13.97mm (0.56in)
    Weight 75gm (2.64oz)

EVALUATION RESULTS

System design and laboratory validation are detailed in cited references[3,4,5]. An integrated system, including accelerometer sensor, was installed inside a steel projectile canister which was fired from a 155-mm howitzer into two 0.3-m (1-ft) thick, steel reinforced, high-strength concrete target walls separated 10.7-cm (0.35 ft). The HDAS system

performed successfully at impact deceleration levels as high as
150,000-g's (the highest shock level tested to date)[4,6], Figure 4.
As many as 10 repeat tests were conducted with the same HDAS module
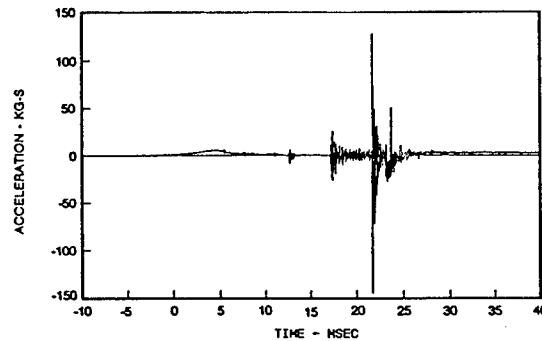without failure.

Figure 4. HDAS-recorded acceleration
wave form from 155-mm
howitzer, "hard" wall
target impact test.

Recent tests were conducted at WES, using the Geomechanics
Division's newly developed Penetration Test Facility[7, 8]. A super high-
strength steel projectile was fired into a nominal 41-Mpa (6-ksi) concrete
(unreinforced) reaction mass. Outstanding data were obtained with an HDAS
system contained within the projectile (Figure 5). The HDAS was sensitive
enough to record the high impact deceleration as well as the much lower
level impacts from the trigger wire and deflector rod (the deflector rod
effected a tilt of 3-degrees in the projectile's exit trajectory).
Figure 6 is a expanded portion of the recorded data wave form detailing
these impacts. Distance calculations derived from the recorded data
matched the physical spacing of the trigger points reasonably well.
Digital low-pass (3-kHz) filtering of the recorded particle acceleration
yielded the lower frequency rigid body acceleration component, Figure 7.
Numerical computer code calculations of rigid body acceleration and
velocity compared well with the recorded data (acceleration, and the
derived velocity), Figures 8 and 9, respectively. Numerical code
calculations were terminated 0.4-msec after projectile impact which
corresponded with the time of rebound of the projectile off the target
wall. Data and calculations at times greater than 0.4-msec after impact
are uninterpretable because the projectile rebounded, rather than
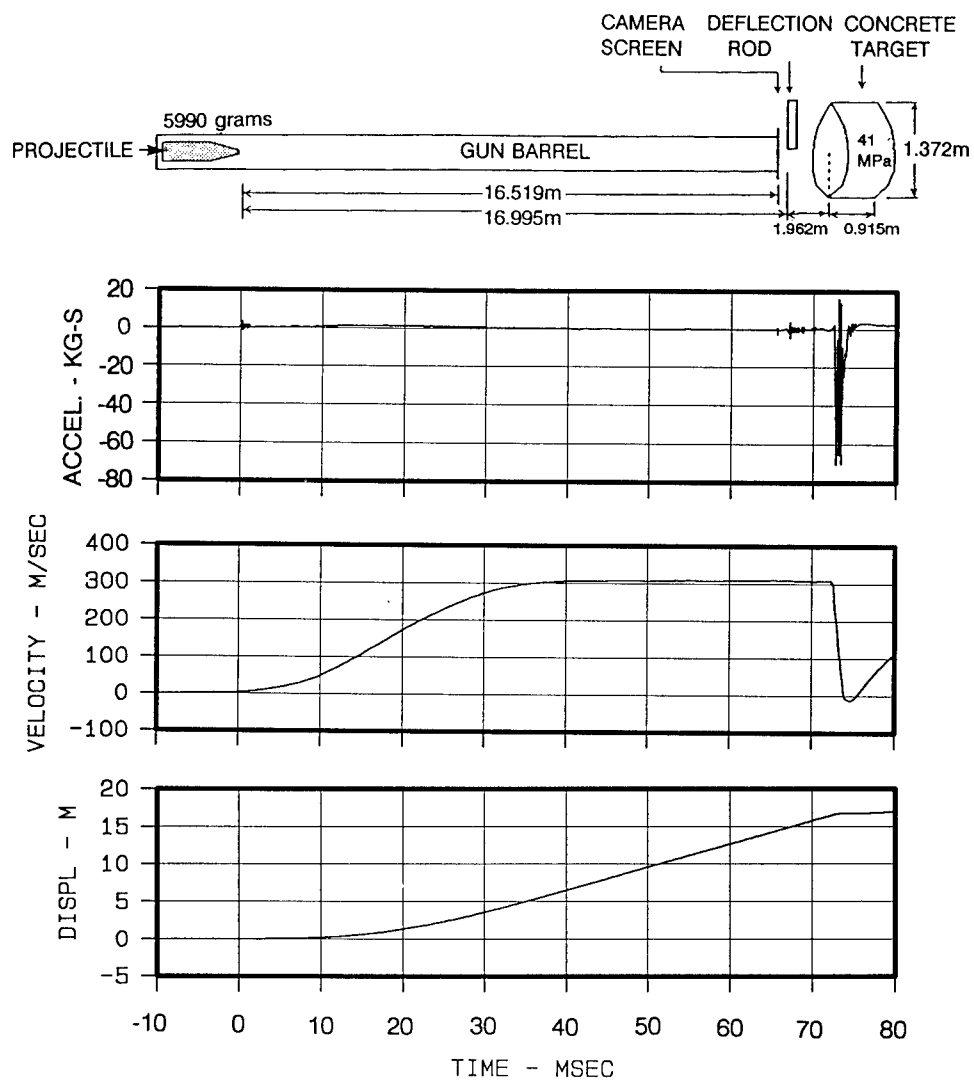penetrated.

Figure 5. Recorded acceleration and derived velocity
and displacement records, WES penetration
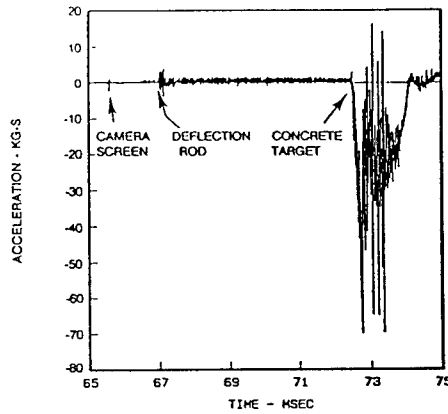tests.

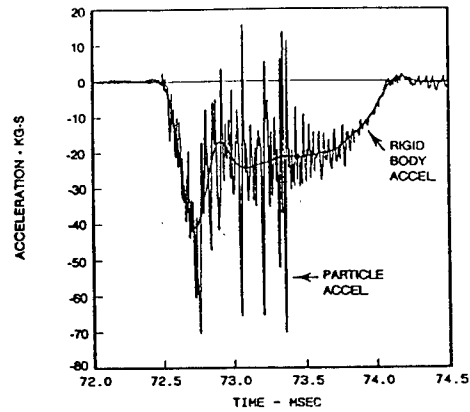Figure 6. Impacts with camera screen, deflection rod, and concrete target.

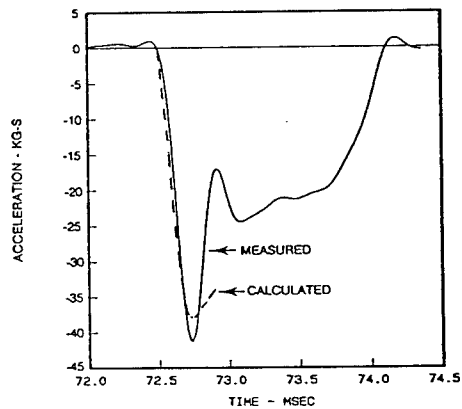Figure 7. Measured particle and rigid-body accelerations.

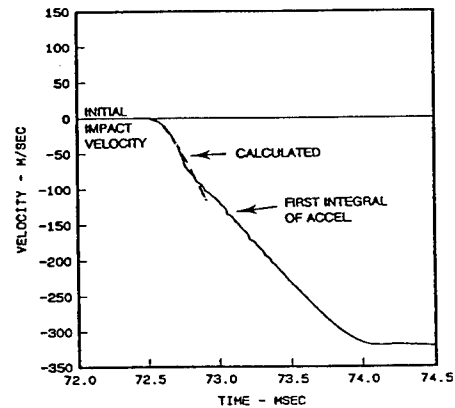Figure 8. Rigid-body acceleration; measured versus calculated.

Figure 9. Velocity comparison; acceleration-derived versus calculated.

625

HDAS systems have been used successfully to record primary blast effects data on two major high-explosive field tests. The first was an in-tunnel detonation of 20,000-kg (44,000-lb) of explosive[5,6,9]. Two HDAS systems were installed inside the mouth of the tunnel and successfully measured high-level dynamic surface airblast pressures, (Figure 10). This is the first time (to our knowledge) that the chamber venting pressure (second, long duration pressure pulse) has ever been measured experimentally. This data provides a crucial empirical reference for validating hydrocode calculations currently being used to predict in-tunnel blast phenomenology. Similarly located airblast gages connected through a conventional hard signal wire system produced only partial, very noisy data, due to early cable failure.
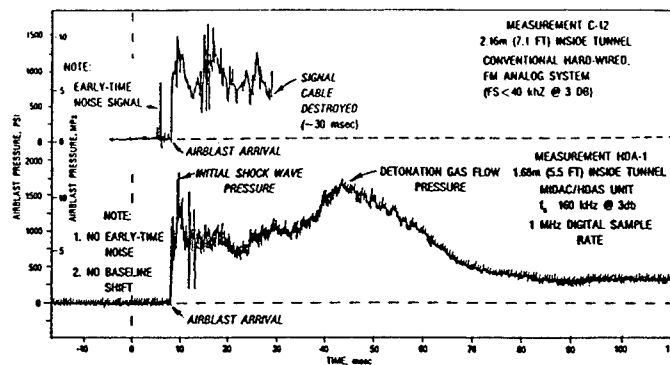
Figure 10. Inside tunnel airblast pressure wave form comparison, HDAS module versus conventional, cabled system (full-scale high-explosive tunnel test).

The second test used three HDAS systems with accelerometers buried in soil near a buried 455-kg (1,000-lb) explosive charge[10], (Figure 11). Although the recorded peak accelerations were relatively low as compared to the maximum capability of the HDAS, this illustrates some of the versatility of the system. The HDAS is flexible enough to operate in practically any environment and at shock levels from seismic to explosive with equal performance. Only the sensor has to be changed out. The sampling rate of the ADC is user-selectable between the maximum rate of 1,000,000-samples per second down to approximately 100,000-samples per second. This provides the user with an effective memory expansion by a factor of 10 when measurement of low frequency phenomena is required.
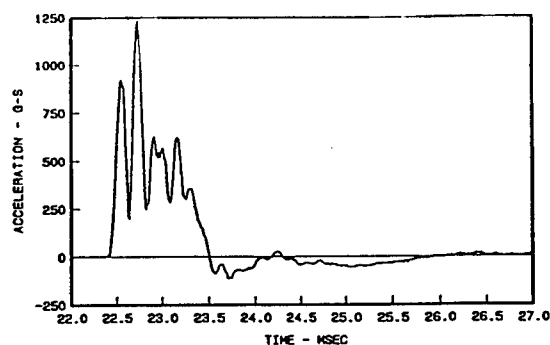
Figure 11. HDAS-recorded soil acceleration from buried
high-explosive detonation.

PERFORMANCE COMPARISON

HDAS performance was compared with that of a commercial "soft",
state-of-the-art digital recorder system. Matched accelerometers were
placed inside a protective canister which was buried in the soil near an
explosive charge. The output signal from one accelerometer was fed to a
commercial digital recorder system located in an airconditioned instrument
van, approximately 61-m (300-ft) from the sensor. Output from the other
accelerometer was fed to a separate canister containing an HDAS unit,
approximately 0.3-m (1-ft) away. Recorded acceleration wave forms from
both systems are compared in Figure 12 and 13. The acceleration wave
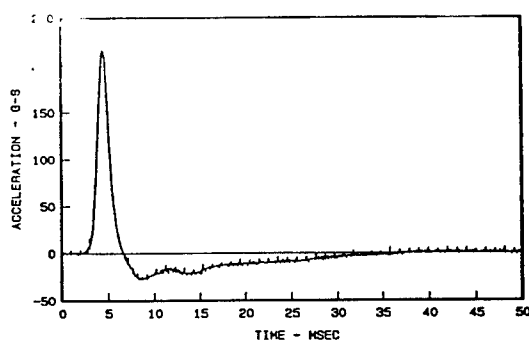
Figure 12. HDAS-recorded
acceleration.
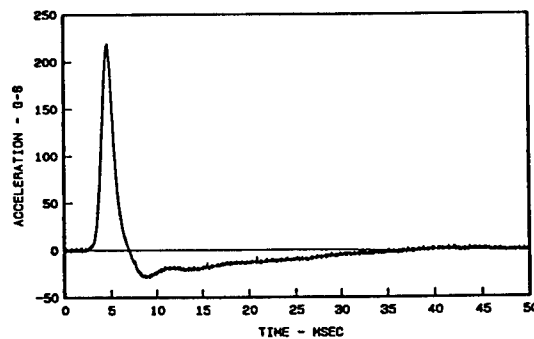
Figure 13. Conventional system -
recorded acceleration.

forms overlay in almost perfect register. An even more telling comparison
is that of the derived velocity and displacement wave forms, Figures 14
and 15. Slight differences in the acceleration wave form are magnified
significantly by the integration process and the length of time over which
the integration is run. The slight differences observable are minimal,
and show outstanding comparison between the two acquisition/ recording
systems, although the commercial system had 12-bit resolution versus
11-bit for the HDAS.
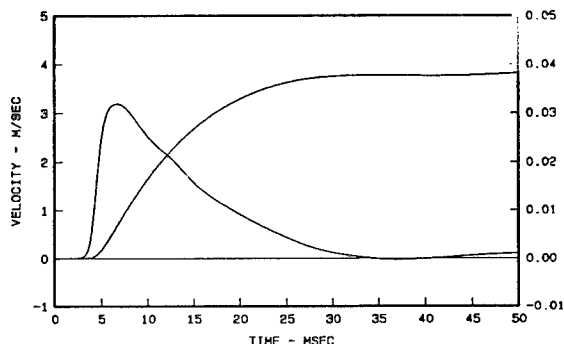
Figure 14. HDAS-derived velocity
and displacement.
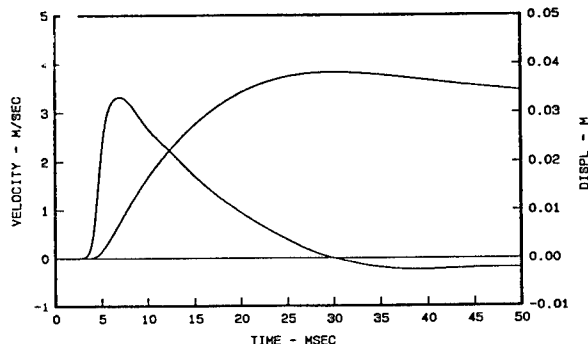
Figure 15. Conventional system -
derived velocity and
displacement.

CONCLUSIONS

An extremely rugged, miniature, high fidelity, stand-alone digital
data acquisition system has been successfully designed, fabricated and
field validated. This HDAS system meets all, and exceeds many, of the
original objectives established for this research program. The current
model has a flat frequency response to 100-kHz, 1-usec sampling (one-
million samples per second), 11-bit resolution, 128-kword on-board memory,
and features such as sensor shunt calibration and synchronization for
multi-channel operation. This versitle device can be used to measure
virtually any dynamic phenomena of interest, from low-level sound pressure
and seismic motion to very high-level impact- and explosion-produced
pressures, stresses, strains, and motions, simply by adjusting the
sampling rate and adding the proper sensor. A significant number of HDAS
units are now in the WES data acquisition systems inventory. An
international high-explosive field test is scheduled to use HDAS
exclusively for data acquisition.

ACKNOWLEDGEMENT

REFERENCES

1. Erdman, Robert V.; "Shock Hardened Data Acquisition System;" Eglin Technical Report AFATL-TR-89-105, June 1989, U. S. Air Force Armament Laboratory, Eglin Air Force Base, FL.

2. Franco, R. J.; "The Small Simple Penetrator 1986 (SSP86) High-G Data Recorder;" SAND89-0873-UC-70F, 6 June 1989, Sandia National Laboratories, Albuquerque, NM.

3. Franco, Raphael A., Jr., "A Self-Contained Shock Hardened (100 kg) Data Acquisition System, Report 1, Design and Development;"WES Technical Report TR 0-88-1, June 1988 US Army Engineer Waterways Experiment Station, CE, Vicksburg, MS.

4. Franco, Raphael A., Jr. and Ingram, James K.; "HDAS, A Miniature, Self-Contained, Super-Hardened, Digital Data Acquisition System (Module);" Paper No. 78, DNA Conference on Instrumentation for Nuclear Weapons Effects Testing (INWET'89), 5-8 June 1989, Monterey, CA.

5. Ingram, James K. and Franco, Raphael A., Jr.; "A Super-Hardened, Self-Contained Digital Data Acquisition System;" 6-10 May 1990, 36th International Instrumentation Symposium, Instrumentation Society of America, Denver, CO.

6. Ingram, James K. and Franco, Raphael A., Jr.; "MIDAC A Miniature Integral Data Acquisition Canister for Blast Pressure Measurement Using the WES-Developed HDAS Autonomous Digital Data Acquisition System (Module);" Paper No. 79, DNA Conference on Instrumentation for Nuclear Weapons Effects Testing (INWET'89), 5-8 June 1989, Monterey, CA.

7.   Rohani, B. and Tidwell, L. A.; <u>Projectile Penetration High-Pressure Soil Test Facility</u>; Army R,D & A Bulletin No. PB70-89-1, January-February, 1990, U. S. Army Material Command, Alexandria, VA.

8.   Franco, Raphael A., Jr.; Results of Penetrator Projectile Impact Tests Using HDAS; Internal Memorandum, January 1990, Instrumentation Services Division, US Army Engineer Waterways Experiment Station, CE, Vicksburg, MS.

9.   Joachim, Charles E.; "Shallow Underground Tunnel/Chamber Explosion Test, Data Analysis;" WES Technical Report (In Publication), US Army Engineer Waterways Experiment Station, CE, Vicksburg, MS.

10.   Unpublished Test Data; MIDNIGHT-HOUR-2 HE Test Event; November 1989, US Army Engineer Waterways Experiment Station, CE, Vicksburg, MS.

FRIEDL, VOGEL & JONES

Assessment of Body Weight Standards in Army Recruits (U)

*Karl E. Friedl, CPT(P), James A. Vogel, Dr., Bruce H. Jones, LTC
Exercise Physiology Division
U.S. Army Research Institute of Environmental Medicine
Natick, MA 07160-5007

INTRODUCTION

Since the early 1980's, the Army has placed increased emphasis on a trim military appearance. This policy has been implemented and rigidly enforced through an expanded program which has become integral to the Army, from mandatory biannual weigh-ins of all personnel to the display of weight and height on personnel efficiency reports. This emphasis is a result of the leadership's conviction that a trim military appearance indicates a disciplined, motivated and combat-ready soldier. The program has effectively eliminated gross obesity from the Army and soldiers have been driven to exercise more and to be more careful about their eating habits. However, these positive aspects of the program are undermined by a mismatched accession regulation which continues to inject fat soldiers into the system and by the appearance of an overly stringent standard for females relative to the male standards.

There are two separately developed and unconnected Army Regulations that bear on weight control, one which allows overfat soldiers into the Army (AR 40-501, Medical Fitness Standards) where they will be in violation of the second (AR 600-9, The Army Weight Control Program). This occurs primarily because the accession regulation uses height-weight tables to identify overfat candidates while the Army Weight Control Program has evolved to a more direct estimation of body fat.[1] This also occurs because the accession weight tables have never been linked to the screening weight tables for the retention standards, and even the age categories are different between the two regulations.

Added to these problems is the appearance of an overly stringent standard for female accessions relative to the male standards. The average body mass index of young men and women in the national population is the same, approximately 22.3 kg/m^2; however, the upper limit of body mass index accepted by the male accession weight standards is 30.9 kg/m^2, while females

FRIEDL, VOGEL & JONES

17-20 YEAR OLD MALES

17-20 YEAR OLD FEMALES

age ranges for male accession standards
and for NHANESII mean values: 16-20

age range for accession standards: 18-20
age range for NHANESII mean value: 16-20
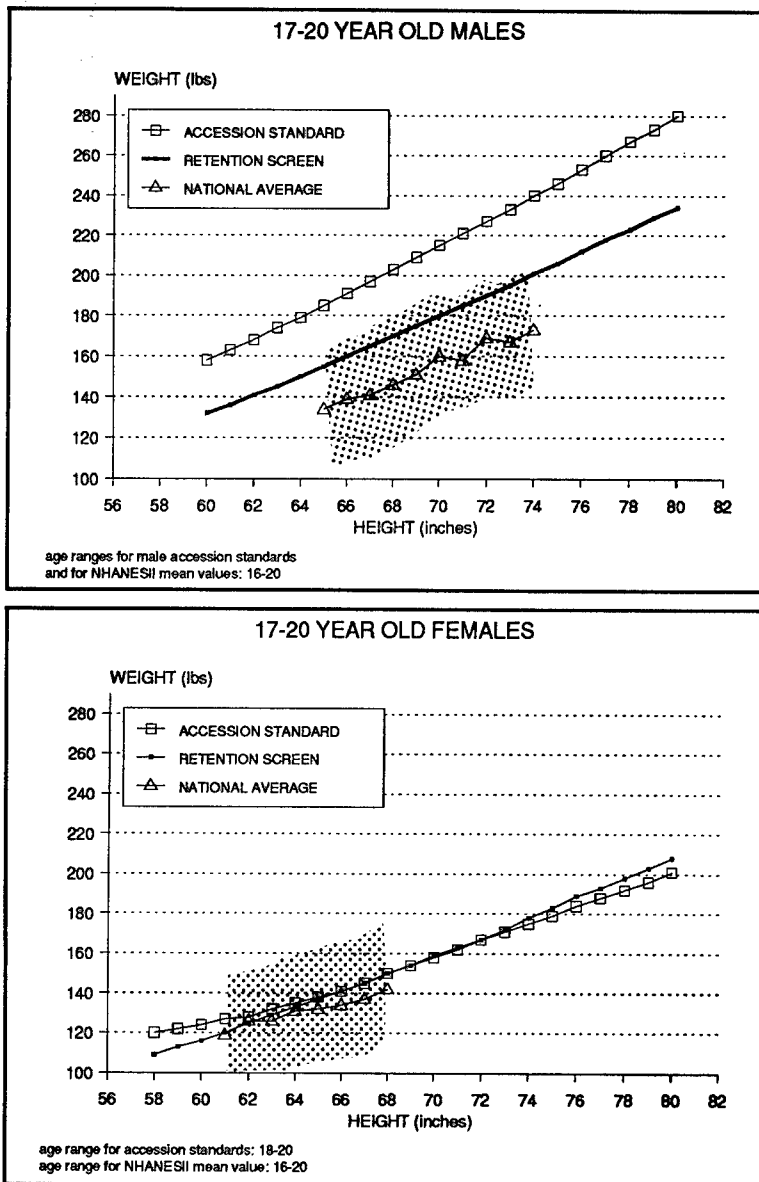
Figure 1. Male (top) and female (bottom) weight for height standards and averages (ages 17-20).

are limited at approximately 22.8 kg/m² (Figure 1). The effect of this difference is that 32% of young women in the national population are excluded from Army enlistment by their weight, while less than 5% of males are similarly excluded.[2] The retention standards are also more restrictive for females compared to males, as reflected in the screening weight standards which precede body fat assessment, but this gender difference is further accentuated by the absence of a weight allowance between the female accession and female retention weight screen while males can exceed the retention weight screen by greater than 40 pounds (see Figure 1).

This study was conducted to explore the relationship between these two standards for males and females. We report here the effect of the accession weight standards on new recruit fatness with respect to retention fat standards, and the relationship of excess fatness, as defined by current retention fat standards, to achieved weight and fat loss in new male and female recruits.

FRIEDL, VOGEL & JONES

METHODS

The purposes and procedures of this study were explained to approximately 2000 new recruits reporting to Fort Jackson for basic training and signed consent was obtained from those who agreed to participate (1894 soldiers). Subjects were recruited from all female companies formed during the study (approximately 1 company/week). Participation was invited for all male soldiers in one or more companies per week, selected from a larger pool of subjects in order to roughly match the number of women in the study. Initial measurements were obtained between 11 September and 9 October 1988. The last group of basic trainees was followed until the end of training (16 December 1988). A mailed followup study of all non-attritted active duty subjects was performed approximately six months after the end of basic training.

The mean age of these recruits was 20.1 ±3.3 (range 17-40) and 20.2 ±3.5 (range 17-35) years for males and females, respectively. Distribution of the three principal ethnic groups represented was: for males - 58.7% white, 30.9% black, and 6.5% hispanic; and for females - 46.8% white, 44.6% black, and 6.2% hispanic. National Guard and Army Reserve Components comprised 20.4% of male and 22.8% of female basic trainees. Female recruits were assigned primarily to low aerobic demand specialties;[3] three fourths were assigned to: 76Y (supply), 94B (cooks), or CMF 33 (signal specialties). No concentration of specialties was evident for male recruits. Detailed analyses of the diet (August 1988) and physical activity (in a 1984 study) of Fort Jackson basic trainees are available in two previous USARIEM technical reports.[4,5]

Height, weight, and circumferences for body fat estimation were measured by a study team of trained observers at the reception station. Height was measured without shoes. Weight was obtained and recorded for soldiers in stockinged feet, t-shirt, and either jeans or BDU trousers and belt. Circumferences were measured with a Gulick tape measure over bare skin, but otherwise in accordance with procedures outlined in AR 600-9. Weight was obtained at the end of basic training from unit records. Height, weight, and circumference measurements were again measured 6 months after basic training, obtained by surveys mailed to company commanders at the soldier's current duty station. These addresses were obtained from the Army Enlisted Masterfile. The printout from this file was performed at an average of 6.7 months after soldiers' completion of basic training. The attrition at the time of the survey was 12.0% (males) and 36.2% (females); with survey data included, overall attrition reached 14.8% (males) and 41.5% (females) of active duty soldiers. An initial mailing to 100 of the study soldiers was performed to test the questionnaire and the mailing list (using an earlier printout). Commanders were given a 20 day suspense from the date of mailing. A followup mailing to all nonrespondents (using the same address list) was made 20 days after the

FRIEDL, VOGEL & JONES

suspense for the first mailout. All returned surveys were analyzed 60 days after the last mailout. Surveys were returned by 75% of the soldiers' unit commanders. Pregnant soldiers were excluded from all analyses (9.8% of returned surveys for female participants).

Data was analyzed with the SPSS-X statistical package (Chicago, ILL), using primarily chi-squared analyses and t-test comparisons. Mean values are shown ±standard deviations. Body mass index (wt/ht^2) was calculated for males and females in this study as the most appropriate unitary expression relating body proportion and size to fatness.[6] The primary breakdown of data in this study was performed as a dichotomous "within" or "exceed" weight-for-height tables, or "within" or "exceed" fat standards.

Percent body fat was calculated according to the relationships listed below (in lbs and inches); these yield the same results as the current tables in AR 600-9, except those measurements are rounded to the nearest 1/4 inch.

Males: % body fat = 46.89 - (68.68 * log (height)) + (76.46 * log (abdominal circumference - neck circumference))

Females: % body fat = (0.44 * hip circumference) + (105.33 * log (weight)) - (1.31 * height) - (3.99 * forearm circumference) - (1.35 * neck circumference) - (0.51 * wrist circumference) - 71.76

Body fat standards for 17-20 year olds are 20% (males) and 28% (females), and for 21-27 year olds, 22% (males) and 30% (females). These two age groups encompass 95% of the new recruits.

The reliability of survey data was tested using team-measured height as a check measurement. Height measurements by the soldiers' first units were overestimated by 0.43 inches. A 0.4 inch (1.0 cm) overestimate in height reduces calculated body fat in the midrange of statures by 0.2% and 0.7% body fat units for men and women. To reduce the influence of this measurement error, all body fat calculations were made using the study team measured height.

RESULTS

Distribution of New Recruits by Weight and Fat Standards

A small percentage of men (5.5%) exceeded even the lenient male accession weight tables. A much larger proportion of these men exceeded the retention weight screen (30.8%), reflecting the large gap between the accession and retention weight tables for males. Female recruits did not range higher above their accession weight standards than males above the male accession

FRIEDL, VOGEL & JONES

weight standards, but 32.1% of female recruits exceeded the accession weight standard and 36.9% exceeded the retention screening weight tables. (One BMI unit corresponds to roughly 6-7 pounds near retention screening limits).

Mean body fat was 16.1 ±5.8% for all males and 26.8 ±4.2% for all females. Males averaged -4.6 ±5.9% body fat units below their limit, while females averaged -1.8 ±4.2%. Broken out by the principal ethnic groups, black recruits were less likely to be overfat than white and hispanic recruits, and both black males and black females were further below their upper limit of fatness than non-blacks (Duncan's multiple range test, p<0.001). Black, white, and hispanic males averaged -6.8 ±5.2, -3.6 ±5.7, and -3.2 ±6.1, and females were -2.6 ±4.0, -1.4 ±3.6, and -0.7 ±3.3% body fat units below their age-specific

Figure 2. Male (top) and female (bottom) distributions around the body fat standards.

standards. Black females, but not black males, were also further below their weight standards.

The distribution of new recruits with respect to the proximity to their retention body fat standards is shown in Figure 2. Essentially all (99.6%) males who were overweight by accession weight standards were also overfat by the
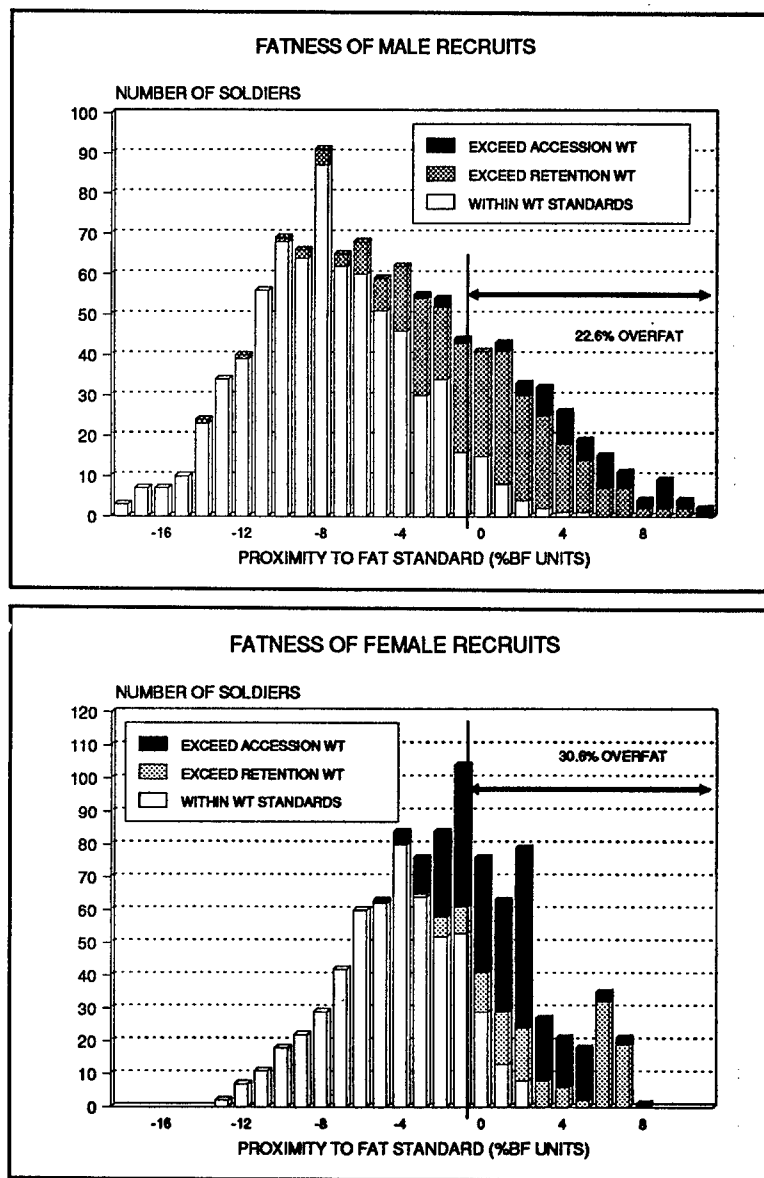
FRIEDL, VOGEL & JONES

retention standards. A large portion of the males who fell below accession weight standards, but who exceeded retention weight screen limits, met body fat standards. The highest body fat measurement in a male recruit within the accession weight standards was 34%, or approximately 14% body fat units over his retention standard.

A portion of females who exceeded accession weight standards were within body fat retention standards (10.9%). Theoretically, these women should not have been allowed into the Army, by the directives in force at the time of this study. Presumably, others in this category were turned away by recruiters for failing to meet the accession weight standards. The distribution of females around their fat standards represents a narrower range than the distribution of men. This is at least partly explained by female body fat standards which come close to the mean value and substantially truncate the right half of the distribution, while the left halves of the male and female distributions are similar.

The minimum portion of new recruits who would qualify for the Army Weight Control Program, if they were immediately held to the regulation, are depicted in the shaded (or solid) areas which span the overfat region. 19.7% of all new male recruits and 25.4% of the female recruits fall into this category.

Weight Loss and Achieved Change in Body Fat

Weight loss during basic training was greatest in the fattest recruits. Mean change in weight compared by initial (entry to active duty, "EAD") fatness is shown for males and females in Figure 3. Females lost weight at all intervals of fatness but did not lose as much weight as males. Overall changes were -0.4 $\pm$6.8 lbs (males) and -2.5 $\pm$5.7 lbs (females) and this weight loss was greatest for the overfat male (-5.1 $\pm$6.4 lbs) and female (-3.2 $\pm$4.1 lbs) soldiers.

Following basic training, male weights continued to change in the same direction taken in basic training, while females at all fatness intervals regained lost weight and added more weight. Overall changes (from EAD to 6 months after basic training) were +1.9 $\pm$11.9 lbs (males) and +4.8 $\pm$7.9 lbs (females). The leanest females lost the least amount of weight during basic training and gained more weight at their first unit, compared to females at higher levels of fatness.

Fat reduction to a within standards mean body fat occurs for male soldiers who are up to 4% body fat above retention standards at EAD. The proportion of overfat males decreased from 23.1% (EAD) to 13.3% (six months) ($p<0.01$). Units reported that 5.8% of all the studied males were on the Army Weight Control Program. By actual count, 46 of 87 (53%) soldiers who were overfat at EAD met their fat standards when surveyed six months after basic training.

FRIEDL, VOGEL & JONES

Females showed changes in both directions, with 12.6% of within- standards soldiers later exceeding fat limits and 35% of initially overfat soldiers later meeting fat standards. The net result of these changes was no significant change in the proportion of overfat females; in this subsample of complete data for retained females 35.4% exceeded fat standards at EAD and 30.7% exceeded standards 6 months after basic training. The proportion of initially overfat females who later met their fat standard was significantly less than for initially overfat males (35% of overfat females vs 53% of overfat males, met standards later; chi squared test, p = 0.02). Units reported that 8.9% of the non-pregnant female soldiers were on the Army Weight Control Program 6 months after basic training.

Figure 3. Weight changes in male (top) and female (bottom) recruits by initial fatness, during basic training, and during basic training plus six months at their first duty station.

DISCUSSION

A primary purpose of accession weight tables is to screen out individuals who are unlikely to be successful in the Army. This lack of success includes the inability to meet retention fat standards. The current female retention and
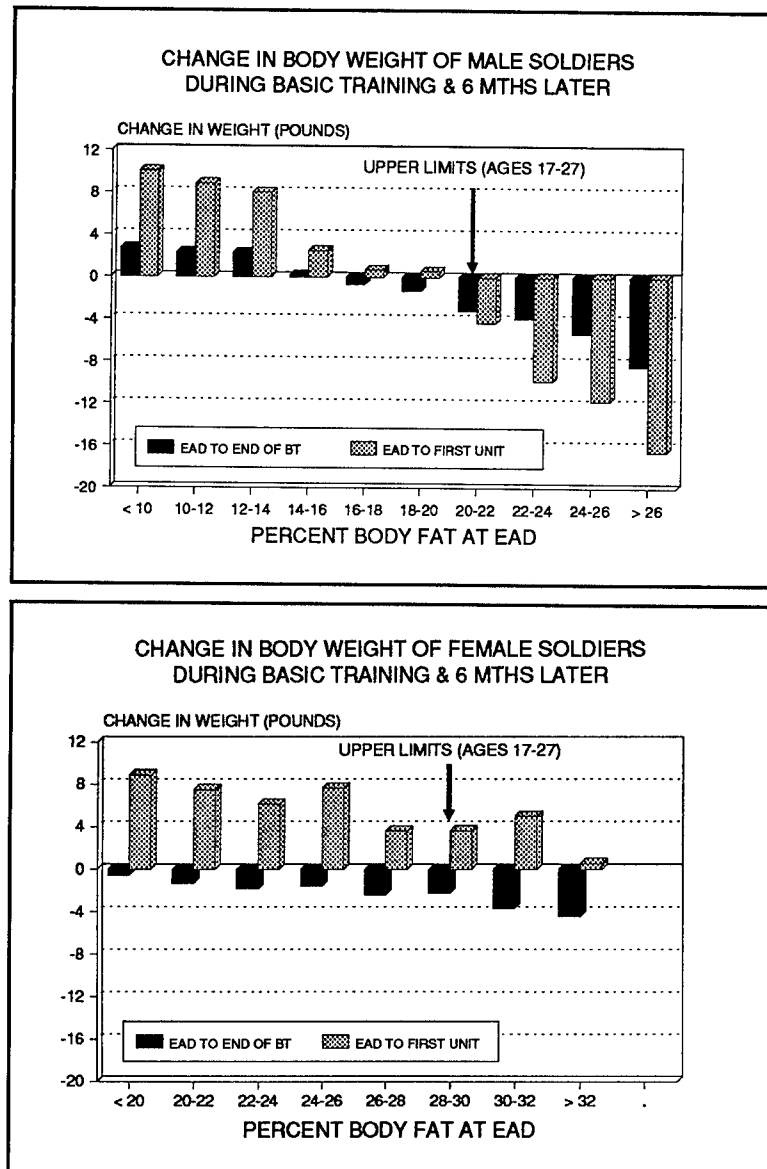
FRIEDL, VOGEL & JONES

accession standards virtually assure that meeting fat standards will be a problem for many females. Most young females in the U.S. population are clustered at the retention fat standards, i.e. the standard is only 1.8% body fat units above the mean fatness of new recruits (a difference which is within the margin of error of most methods of body fat estimation) while males have a more generous safety zone between mean fatness and the upper limit allowed. The restrictive female retention standard results in a considerable spillover into the overfat range. Thus, even though the female accession weight standards are not more lenient than the retention weight screening tables, 30.6% of all new female recruits are overfat. Even so classified, most of these recruits only slightly exceed the retention standards and are not fat by comparison to national averages. One fourth of all female recruits exceeded the standard by less than 4% body fat units (equivalent to 5 or 6 pounds of fat weight in an average female recruit).

In contrast, the current accession weight tables permit entry of virtually all otherwise qualified young males, including some very fat males. However, the proportion of overfat male recruits who are accepted under these lenient weight standards is still smaller (23.1%) compared to the proportion of overfat females. This is due to the fact that most young males in the U.S. population are considerably below the Army fat standards. Compared to female recruits, very few males exceeded accession weight standards and were within fat standards (<0.5%).

The various standards and averages can be better appreciated if the same property, body fat, is compared. Body fat standards are used for individual assessments in the Army Weight Control Program because they more directly reflect obesity than body size (weight-for-height), but the two are correlated and an average fatness can be predicted for a group of recruits of given body mass index. Accordingly, Table 1, below, shows percent body fat equivalents estimated from body mass index using regressions from our current Army recruit sample. This comparison shows how close the female retention fat standards are to the estimated fatness of female in a national sample, while male national means are well below Army limits. The difference is even greater when compared to accession tables. Thus, there is a sizeable allowance for male fatness between the accession and retention standards, and again between the retention standards and the national average. There are no such differences for females, where the approximated body fats are virtually identical between the national average, the retention fat limit, and the accession fat limit equivalent.

Body fat/body weight standards set solely on the basis of health considerations would greatly increase female fatness. It should be noted that percent body fat increases faster than body mass index, thus 20% over the "desirable" weight-for-height (the approximate upper limit for health

FRIEDL, VOGEL & JONES

considerations prescribed by the Surgeon General, Public Health Service) would be reflected as a 10% body fat unit increase over estimated national means, or a 60% increase in fatness of young males and a 33% increase in fatness of young females. Since the objectives of the Army Weight Control Program are to ensure a minimum level of physical performance and military appearance, these health standards would be unacceptable because they are far above fatness thresholds associated with acceptable military appearance and aerobic fitness.[7]

Observed changes in weight and circumference-based fatness provide some rationale for a continued difference in allowances between retention and accession standards for males and females. Males who exceeded their retention standard by up to 4% body fat were likely to achieve their fat standard by six months after basic training. This suggests that males, young males, as an example, could be allowed 24% for accession and be expected to meet the current retention standard of 20% body fat within a six month period of time after basic training. There is no comparable justification for such an allowance based on an achieveable fat loss in females. Although some overfat females later achieved their standards, a larger proportion of overfat females, compared to overfat males, did not. On average, female recruits at their first unit were at the same level of fatness as when they first entered to active duty, even though some transient weight (and presumed fat) loss was accomplished in the interim period of basic training.

Table 1. Percent body fat of recruits and retention fat standards compared to equivalent values of percent body fat for the current accession weight standards and for the national average (estimated from body mass index).

Age category -->	Males			Females		
	17-20	21-27	28-39	17-20	21-27	28-39
Recruits in this study (mean value)	15.6	17.0		26.8	26.3	
*NHANESII weight (mean value)	(15)	(19)	(20)	(27)	(30)	(33)
Retention fat standards (upper limit)	20	22	24	28	30	32
*Accession weight tables (upper limit)	(32)	(34)	(33)	(28)	(30)	(32)

* estimated from regression of BMI and %body fat estimated by AR 600-9, from recruits in this study. Equations: males: %body fat=(BMI-14.67)/0.514, r=0.82, n=1048; females: %body fat=(BMI-9.89)/0.459, r=0.86, n=816.

FRIEDL, VOGEL & JONES

These findings are derived from the use of the current Army circumference equations for predicting body fat. To date, these equations have not been tested for suitability in evaluating small and/or acute changes in body fat as employed here. Thus, their suitability for the accurate assessment of body fat change is unknown. Furthermore, the separately developed male and female equations, using different measurements, may not estimate body fat changes to the same degree of accuracy in both genders. Despite these limitations, other evidence suggests that the gender differences in achievement of fat standards after basic training are real. Females almost certainly lost body fat during basic training, since at all levels of fatness there was a mean reduction in weight, even while lean body mass may have been increasing.[8] Following basic training, there was a mean increase in weight at all intervals of fatness. This can be reasonably assumed to be due to fat rather than lean body mass gain, since exercise levels for most soldiers would have been greatest during basic training, not after basic training. This is also consistent with a previous study from this Institute which followed a small sample of West Point cadets over a two year period finding an initial reduction in body fat when the training was intensive, and then a gradual increase to higher than initial fatness.[9] In the current study, hip circumference (weighted against "fat" in the female equation) increased by more than 1/2" between EAD and 6 months after basic training. Along with the average 5 lb gain in weight, these two factors represent a theoretical increase of approximately 2% body fat units for the typical 125 lb female recruit. These estimations are offset by increases in neck, arm, and wrist circumferences which represent lean body mass in the equation. Thus, changes in the actual measurements used in the female body fat equation counter the likelihood of a significant mean female fat reduction, even though a transient reduction probably occured during basic training.

Job assignments may provide at least a partial explanation of the weight gain in females following basic training. Nearly all of the assigned job specialties for females in this study group are sedentary, while males were assigned to many jobs which require a high level of aerobic and/or strength capability and involve a high level of regular physical activity. The temporary weight loss in basic training indicates that the longer term outcome might be improved if effective weight loss assistance in terms of exercise and reduced caloric intake was provided to overfat soldiers. Our results suggest that female soldiers fresh from basic training should be prime targets of unit physical fitness programs.

An alternative reason for the gender difference in weight changes and success in achieving fat standards by overfat recruits can also be suggested. Male soldiers may be more highly motivated to achieve their weight/fat standards for retention in the Army because they range further above their standards at entry, compared to most overfat females who are only a few

FRIEDL, VOGEL & JONES

pounds away from their retention screening weights. In other words, if females had also been afforded the wide gap between liberal accession weight standards and their retention standards, a higher proportion of overfat females may have been motivated to achieve weight/fat loss. This requires an assumption that new recruits are aware of the weight/fat standards. The hypothesis is not consistent with data which suggests physiological differences in the ability to mobilize fat from some sites of female fat deposition.[10,11] A recent 10 year followup to the NHANESI data also demonstrates that young females are twice as likely as males to have a major weight gain and that initially overweight females have the highest incidence of major weight gain.[12] Thus, such an experiment would be likely to introduce females into the Army who would be unable to achieve, or later maintain, retention fat standards.

Although a change to body fat accession standards would ensure recruitment of men and women who are better able to comply with the standards of the Army Weight Control Program, this would also substantially change the availability of recruits. With no allowance (i.e. accession standards = retention standards), 22.7% of current male recruits could be excluded. An allowance of 4% over each of the current age standards would exclude 8.5% of the males currently recruited (Table 2). This would substantially narrow the current margin of allowance for males while excluding only the fattest males who are unlikely to achieve retention standards.

The effect of no accession allowance for female recruits is quite drastic because so many are so close to their upper limit. If candidates were held to retention body fat standards, 31.6% of all current female recruits could be excluded (12.6% of the remainder had become overfat at their first unit). This loss of eligible female candidates would be partially offset by an unknown proportion of females who are within fat standards but overweight, who were excluded from this sample of recruits. (Since the time in which our data was collected, recruits are also being accepted if they meet the retention body fat

Table 2. Percent of new male and female recruits affected and total excluded with allowances above retention body fat limits.

%body fat units above standard	% Males in group	Cumulative % excluded	% Females in group	Cumulative % excluded
no allowance	---	22.7	---	31.6
+2% body fat	8.0	14.7	14.6	17.0
+4% body fat	6.2	8.5	11.9	5.1
+6% body fat	4.3	4.2	4.4	0.7

FRIEDL, VOGEL & JONES

standards.) A 4% allowance above the current fat standard for females would exclude only 5.2% of current recruits while continuing to allow at least one fourth of all entering females to be in violation of the fat standards at their first unit.

Clearly, a relaxation of female accession standards alone would undermine the Army Weight Control Program by allowing entry of even more overfat females who appear to be unlikely to achieve and maintain fat standards. There are three possible solutions to this problem of overfat female recruits:

1) simply exclude all overfat females by applying appropriate accession standards based on current retention body fat standards, accepting a one fourth to one third reduction in otherwise qualified female accessions,

2) promote a more effective command-sponsored weight control and fitness program following basic training, with special emphasis on females to ensure that they will achieve and maintain fat standards, and/or

3) instead of unilaterally relaxing accession standards, liberalize female retention standards (AR 600-9) within acceptable allowances of the objectives of the Army Weight Control Program (i.e. military appearance and physical fitness).

Regardless of the approach, it is unreasonable to continue to induct one out of three females into the Army who are not in compliance with the fat standards of AR 600-9, a portion of whom will be eliminated by administrative separation for failure to meet weight control standards, irrespective of specialized training and job performance. It is apparent from this study that the concern about accession standards treating males and females with equal fairness should extend to the retention standards. Specifically, female retention body fat standards should be reconsidered and possibly increased (e.g. at least +2% body fat units) to reflect the same confidence interval between mean values and the upper allowable limit of % body fat which is allowed for males.

CONCLUSIONS

Accepting that the accession standards should be linked to the retention standards by estimates of achievable fat loss, it is also reasonable to suggest that the same system of evaluation should be used for both standards. Thus, a body fat standard along with an initial body weight screen should be used for accession as it is employed for retention. Using body fat instead of body weight as the ultimate accession standard will save the accession of women who are currently excluded as overweight but who are not overfat, and will also help to exclude male and female recruits who are not likely to meet retention

FRIEDL, VOGEL & JONES

standards. Accession fat standards for males should be brought much closer to retention fat standards than the level of fatness allowed by the current accession weight standards, but an allowance of 4% body fat units over retention standards is within the bounds of observed fat loss in new male recruits. Retention fat standards for females should be further evaluated with consideration to liberalizing these limits to reflect a wider confidence interval between mean fatness of suitable female candidates and the upper limit. Military appearance and aerobic performance should be evaluated to consider why there should not be a +2% body fat unit increase in the female body fat standards. Greater command emphasis on fitness maintenance programs for female soldiers should be a priority. This would reflect a more realistic approach to fat standards and lead to a more effective Army Weight Control Program.

REFERENCES

1. Friedl KE (1990). Body composition and military performance: origins of the Army standards. Proceedings, Committee on Military Nutrition Research, National Academy Press. (to be published)
2. Laurence MT (1985). Development of a methodology for establishing joint service height and weight standards for enlistment. Defense Manpower Data Center, Arlington, VA November 1985.
3. Vogel JA, JE Wright, JF Patton, J Dawson, MP Eschenback (1980). A system for establishing occupationally-related gender-free physical fitness standards. No. 5-80, USARIEM, Natick, MA. March 1980.
4. Rose RW, CJ Baker, C Slater, W Wisnaskas, JSA Edwards, MS Rose (1988). Dietary assessment of US Army basic trainees at Fort Jackson, SC. No. 6-89, USARIEM, Natick, MA. Dec 1988.
5. Jones B, R Manikowski, J Harris, J Dziados, S Norton, T Ewart, JA Vogel (1988). Incidence of and risk factors for injury and illness among male and female Army basic trainees. No. 19-88, USARIEM, Natick, MA. June 1988.
6. Garrow JS, J Webster (1985). Quetelet's index (w/h^2) as a measure of fatness. Int J Obes 9:147-153.
7. Friedl KE, JA Vogel (1990). Army body fat standards: the relation to acceptable military appearance and aerobic performance. (in preparation)
8. Vogel JA, JP Crowdy, AF Amor, DE Worsley (1978). Changes in aerobic fitness and body fat during Army recruit training. Eur J Appl Physiol 40:37-43.
9. Daniels WL, JE Wright, DS Sharp, DM Kowal, RP Mello, RS Stauffer (1982). The effect of two years' training on aerobic power and muscle strength in male and female cadets. Aviat Space Environ Med 53:117-21.
10. Rebuffe-Scrive M, P Lonnroth, P Marin, C Wesslau, P Bjorntorp, U Smith

FRIEDL, VOGEL & JONES

(1987). Regional adipose tissue metabolism in men and postmenopausal women. Int J Obes 11:347-355.
11. Fried SK, JG Kral (1987). Sex differences in regional distribution of fat cell size and lipoprotein lipase activity in morbidly obese patients. Int J Obes 11:129-140.
12. Williamson DF, HS Kahn, PL Remington, RF Anda (1990). The 10-year incidence of overweight and major weight gain in U.S. adults. Arch Int Med 150:665-672.

Helicopter Rotor Blade Ballistic Damage Effects
on Helicopter Dynamic Characteristics (U)

J.C.Fries(*)
Aerospace Engineer
US Army Ballistic Research Laboratory
SLCBR-VL-A,APG,Md.21005-5066

Introduction

In the analytical development, battle damage simulations cause variations of the rotor blade stiffness, mass, chordwise center of gravity, and the elastic axis. Variation of the position on the blade where damage may occur is also taken into account, and the effects that these parameters have on the rotor blade's natural vibration characteristics are investigated for blade torsion and flap bending.

The rotating blade's natural vibration frequencies and mode shapes are analyzed by utilizing a transfer matrix computer code developed at BRL. With this analysis, the eigenvalues and eigenvectors are calculated to yield a state vector of deflections, slopes, moments, and shears all along the blade span.

Rotor blade damage is treated by altering the physical properties to reflect battle damage and studying the resultant changes to the natural frequencies and mode shapes.

In addition to the blade's natural characteristics, a forced response analysis has been developed using the natural characteristics in conjunction with the structural dynamics of the rotor blade and the aerodynamic forcing of the rotary wing. The loads produced by the rotor blade against the rotor hub are calculated to predict the shaking vibrations that are transmitted into the cockpit.

Comparisons are made between the cockpit vibration levels felt by a pilot for a normal balanced rotor and a damaged rotor. Tolerance of the pilot to these vibrations is estimated.

I.Analytical Development

Rotor Blade Natural Modes and Frequencies

The general transfer matrix method,reference 1, is used here to treat helicopter rotor blade natural modes and frequencies of vibration. The rotor blade is divided into an arbitrary number of sections. The physical properties over each section is assumed to be constant, i.e. the mass properties of each section is represented as a lumped mass, and the flexural stiffness is treated as a massless uniform beam.

II.Discussion

This paper reports application of the methods developed to the CH-47 Chinook and UH-60A Blackhawk aircraft. The emphasis for the CH-47 is on the natural vibration characteristics, while the UH-60A emphasis is placed on the forced vibration response.

Description of the Damage Mechanisms

Three types of damage mechanisms are treated. First, holes of various diameters at six different positions placed one at a time on one blade of a rotor set are studied. A second damage mode is the partial loss of an outer section of one blade of a rotor set, and the third mechanism is the loss of a pitch link of one rotor blade.

Natural Characteristics (Chinook)

For the Chinook, it is found that hole sizes through the rotor blade's spar with a diameter of at least five inches is required to produce a significant change in the natural frequencies. Therefore, all the hole damage effects reported herein are for hole diameters of this size.

1.Spar Hole Damage Geometry

Figure II.1 illustrates a hole through the spar with a diameter Δ L forward of the center line. This hole is assumed to be caused by a projectile penetration through the top and bottom thicknesses of the spar.
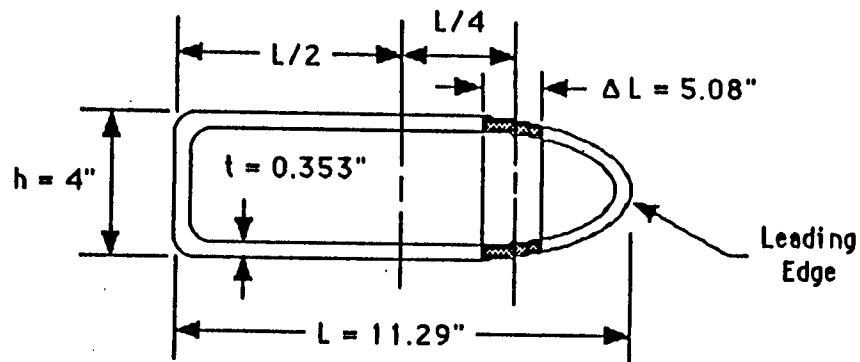

Figure II.1,Spar Hole Damage Geometry

2.Blade Hole Damage Geometry

Figure II.2 shows the hole damage locations in the plan form of the rotor blade studied in this treatment. Six locations are selected at three radial locations.For each location, the hole can be placed either fore or aft of the spar center line.
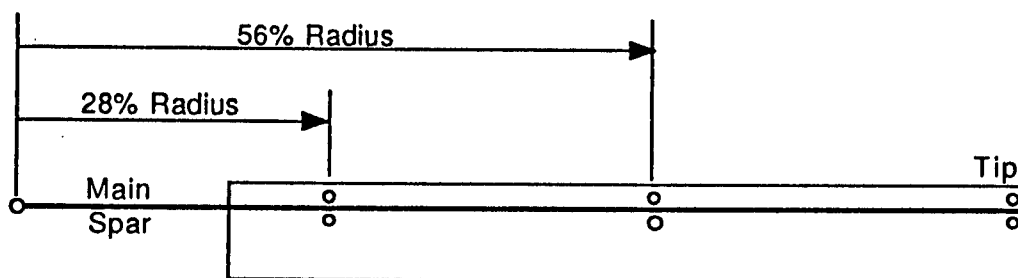
Figure II.2,Blade Hole Damage Geometry

These holes in the spar cause loss of blade weight, reduc-
tion of the local stiffness, and a shift in the local blade
chordwise center of gravity.

3.Torsional Natural Frequency Ratio Variations with Hole Position

The holes through the spar cause the torsional natural
frequencies to shift by the amounts listed in table 1.

Table1

Torsional Natural Frequency Ratio Variations
Due to Hole Position

Hole Position	Natural Frequency Ratio
.28Radius,Fwd	4.58
.28Radius,Aft	4.58
.56Radius,Fwd	4.60
.56Radius,Aft	4.60
Tip,Fwd	4.63
Tip,Aft	4.64

We see here in table 1 that the tendency is for the torsional natural frequency to decrease as the hole moves from the tip to the root of the blade. The natural frequency ratio of the undamaged blade is 4.62 at a rotor speed of 230 RPM, i.e. torsional frequency/230=4.62. Hole damage at the tip causes an increase in the frequency ratio to 4.63 for a forward hole, and 4.64 for an aft hole. As the hole position is moved inboard, the frequency ratio drops to 4.58

4. Torsional Shear Stress Ratio Variations Due to Hole Damage

Calculations have been made to assess the effect that hole damage in the spar has on the shear stress in the blade. Since we are dealing with natural characteristics and natural modes, the modal moments and deflections are relative to an arbitrarily imposed boundary condition. Therefore only relative shear stresses are calculated, i.e. a ratio of shear stresses for the damaged blade relative to the shear stresses for the undamaged rotor blade. The results are shown in table 2. Two separate independent calculations are made and they produce almost identical results. One calculation uses the modal moments as the basis of calculation and the second method utilizes the modal deflections.

Table 2

Torsional Shear Stress Ratio Variations
Due to Hole Position

Blade Station	Shear Stress Ratio Based on Modal Moments	Shear Stress Ratio Based on Modal Deflections
.28Radius	1.3196	1.3183
.56Radius	1.3250	1.3219
Tip,Fore	1.2460	1.2490
Tip,Aft	1.1769	1.1762

It is seen from table 2 that hole damage will increase the local shear stress in the hole region by amounts of 18% at the tip to 32% at the inboard hole positions. In the blade design process, safety factors have to be allowed for the blade to tolerate the increased stress levels due to the damage expected.

5. Torsional Natural Frequency Ratio Variations Due to a Loss of the Blade Tip

The next type of damage that is considered is the loss of a rotor blade's tip section.Figure II.3 shows how the torsional natural frequency ratio changes as a function of the amount of the blade tip that is removed.
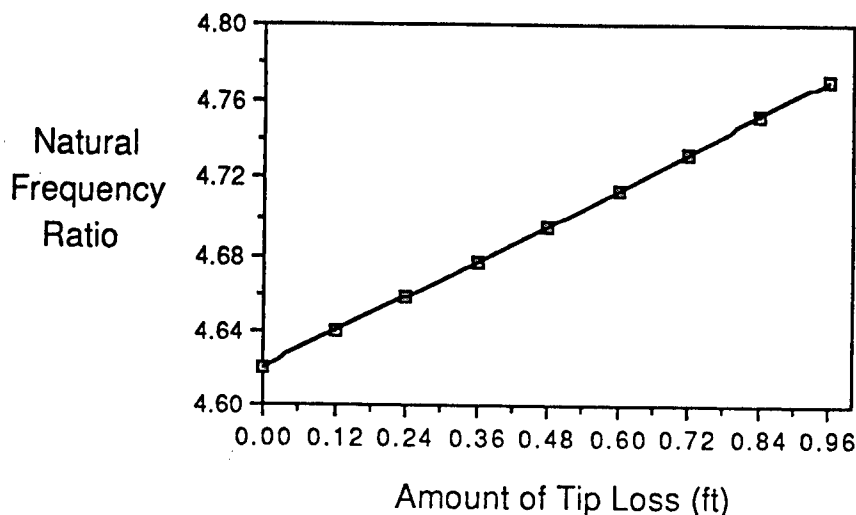
Figure II.3,Torsional Natural Frequency Ratio Variation
with Tip Loss

We see here that, as sections are removed, the natural frequency ratio increases from 4.62 for the undamaged blade to 4.77 when .96 ft is removed from the blade tip.

6. Torsional Shear Stress Ratio Variations Due to a .96 ft Tip Loss

Calculating the shear stress ratios for the case where .96 ft of the blade tip is lost, gives values as presented in table 3.

Table 3

Torsional Shear Stress Ratio Variations
Due to a .96 ft Tip Loss

Blade Station	Shear Stress Ratio Based on Modal Moments	Shear Stress Ratio Based on Modal Deflections
.84Radius	0.8618	0.8813
.88Radius	0.8064	0.8373
.92Radius	0.7075	0.7646
.96Radius	0.4806	0.6204
1.00Radius	0.1992	0.2010

In real dimensional terms, 1.00 in table 3 represents the blade length after removal of .96 ft of the tip. We see from table 3 that the shear stress ratios are reduced all along the blade, but more pronounced at the tip. However, this benefit is offset by rotor inbalance due to having one dissimilar blade in the rotor set.This would cause unwanted vibrations to be transmitted to the fuselage.

7.Torsional Natural Frequency Ratio Variations Due to a Pitch Link Loss

The occurrence of a failed pitch link produces a drastic drop in the torsional natural frequency ratio from 4.62 to 1.0.

8.Torsional Shear Stress Ratio Variations Due to a Pitch Link Failure

The torsional shear stress ratio calculations produce very large stress ratio increases along the rotor blade, amounting to an almost fivefold increase at the outboard end.This is displayed in table 4.

Table 4

Torsional Shear Stress Ratio Variations
Due to a Pitch Link Loss

Blade Station	Shear Stress Ratio Based on Modal Moments	Shear Stress Ratio Based on Modal Deflections
.68Radius	3.879	3.777
.72Radius	4.115	3.997
.76Radius	4.361	4.247
.80Radius	4.594	4.481
.84Radius	4.775	4.720
.88Radius	4.855	4.853
.92Radius	4.848	4.917
1.0Radius	4.741	4.749

9.Flap Bending Natural Frequency Ratios for the Undamaged Blade

In flap bending of the undamaged blade,there are three frequencies of interest.They are as listed in table 5.

Table 5

Flap Bending Natural Frequency Ratio

Mode	Natural Frequency Ratio
1	1.020
2	2.71
3	5.31

The first flap mode is an almost rigid body mode, having very little bending associated with it. This mode is therefore neglected in the analysis. Only the second and third modes are considered.

10. Flap Bending Natural Frequency Ratio Variations Due to Hole Position

Table 6 shows how the flap bending natural frequency ratio changes due to hole damage position.

Table 6

Flap Bending Natural Frequency Ratio Variations Due to Hole Position

Hole Position	2nd Mode	3rd Mode
Tip	2.71	5.33
.58Radius	2.71	5.31
.28Radius	2.71	5.29

In the second mode, no frequency ratio variation occurs, but for the third mode the frequency ratio decreases from 5.33 at the blade tip to 5.29 inboard on the blade.

11. Flap Bending Stress Ratio Variations Due to Hole Position

Tables 7 & 8 show the results of the stress ratio changes with hole position.

Table 7

Flap Bending Stress Ratio Variations Due to Hole Position, 2nd Mode

Blade Station	Stress Ratio Based on Modal Moments	Stress Ratio Based on Modal Slopes
.28Radius	1.5040	1.4924
.56Radius	1.5120	1.5017
Tip	1.5595	1.5629

Table 8

Flap Bending Stress Ratio Variations
Due to Hole Position,3rd Mode

Blade Station	Stress Ratio Based on Modal Moments	Stress Ratio Based on Modal Slopes
.28Radius	1.5100	1.4973
.56Radius	1.5284	1.5185
Tip	1.5255	1.5236

Tables 7 & 8 show that hole damage causes stress ratio inceases up to 56% for the second mode, and 52% in the third mode.

12. Flap Bending Natural Frequency Ratio Variations Due to a Tip Loss

When the rotor blade experiences loss of a tip section, the flap bending frequency ratios change as illustrated in figures II.4 & II.5.From these figures we see the frequency ratios change from 2.71 to 2.78 in the second mode and 5.31 to 5.70 for the 3rd mode when 23 inches of the tip is removed.
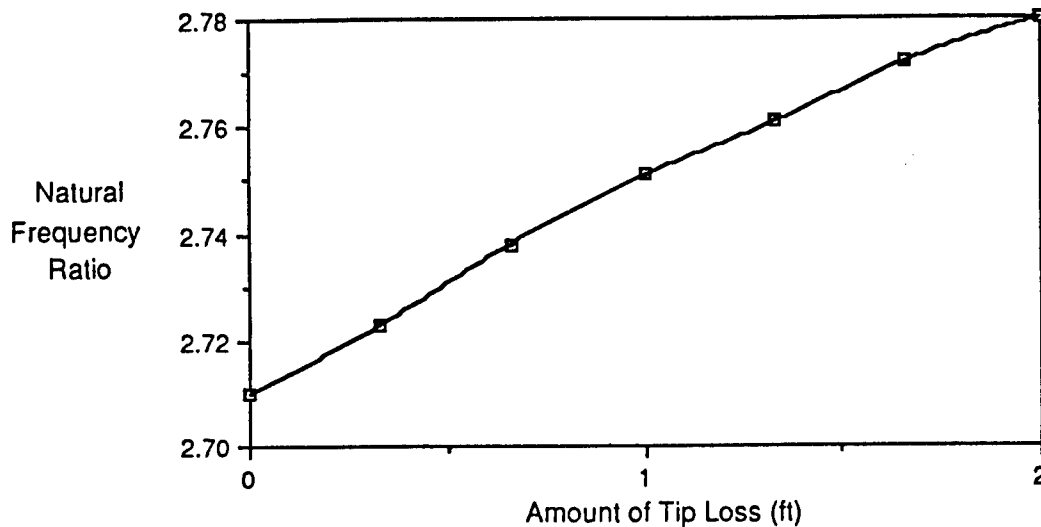

Figure II.4,Flap Bending Natural Frequency Ratio Variation
with Tip Loss,2nd Mode
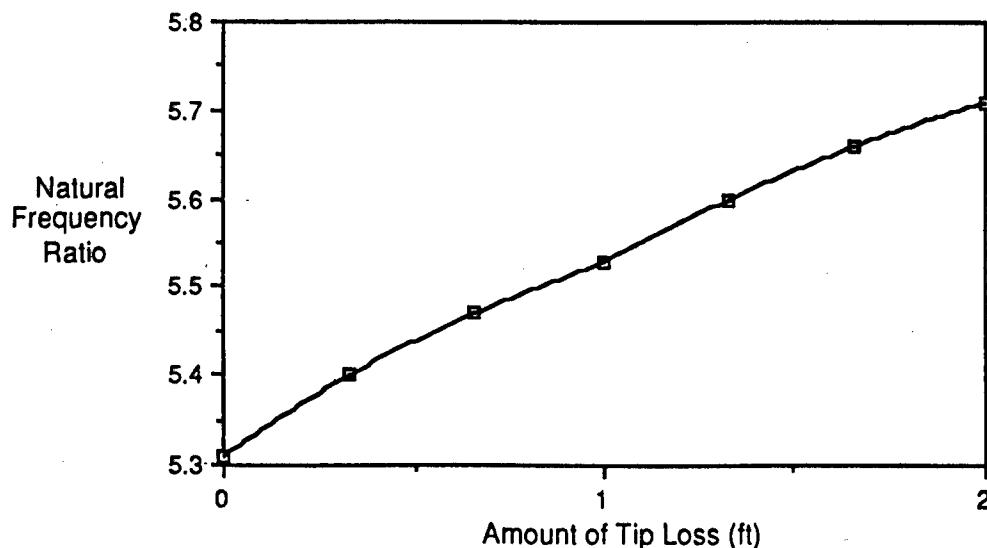
653

Figure II.5,Flap Bending Natural Frequency Ratio Variation
with Tip Loss,3rd Mode

13.Flap Bending Stress Ratio Variations Due to a Blade Tip
Loss

The flap bending stress ratios for modes 2 & 3 for the
case where .96 ft of the tip is lost is shown in tables 9 &
10.

Table 9

Flap Bending Stress Ratio Variations
Due to a .96 ft Tip Loss,2nd Mode

Blade Station	Stress Ratio Based on Modal Moments	Stress Ratio Based on Modal Slopes
Root	1.006	1.006
.022Radius	1.057	1.058
.022Radius	1.394	1.285
.08Radius	1.228	1.333
.12Radius	1.218	1.206
.16Radius	1.215	1.215
.20Radius	1.211	1.215
.24Radius	1.2095	1.211

Table 10

Flap Bending Shear Ratio Variations
Due to a .96 ft Tip Loss,3rd Mode

Blade Station	Stress Ratio Based on Modal Moment	Stress Ratio Based on Modal Slope
Root	0.910	0.997
.022Radius	0.996	0.997
.022Radius	1.197	1.201
.08Radius	1.170	1.170
.12Radius	1.161	1.162
.16Radius	1.150	1.151

From these tables 9 & 10, we see that stress ratio increases occur inboard for both modes.Mode 2 exhibits a 39% increase, and mode 3, a 20% increase.

Forced Response Vibrations (Black Hawk)

As was mentioned earlier, the second helicopter studied here is the UH-60A. This study consists of analyzing the UH-60A helicopter for the effect of main rotor damage on the vibration levels in the cockkpit.

To do this, first an analysis is performed to analyze the Black Hawk's natural rotor blade characteristics in a fashion similar to that performed on the Chinook helicopter.Once the natural characteristics are obtained, then a subsequent computer code is used to calculate the aerodynamic rotor blade forces, the blades' dynamic response, and the vibrations that result.

The same types of blade damages are imposed onto the Black Hawk blade as reported for the Chinook, i.e. hole damage and loss of the blade's tip. However pitch link loss is not reported.The hole diameter damage is 3.37 inches for the Blackhawk, because smaller size holes did not produce a significant change in the dynamic characteristics.

1.Survey of Human Tolerance to Vibrations

In addition to knowing the magnitudes and frequencies of vibrations that undamaged and damaged blades transmit to the cockpit, an evaluation is made to determine how the pilot may tolerate the vibrations.

A literature search yielded data that has been accumulated by a number of investigators.The most useful report is reference 3. From this report, data is extracted and put in a form compatible with the format used in BRL's analysis.

2.Results of Human Tolerance and Forced Vibration Levels Generated by an Undamaged Rotor
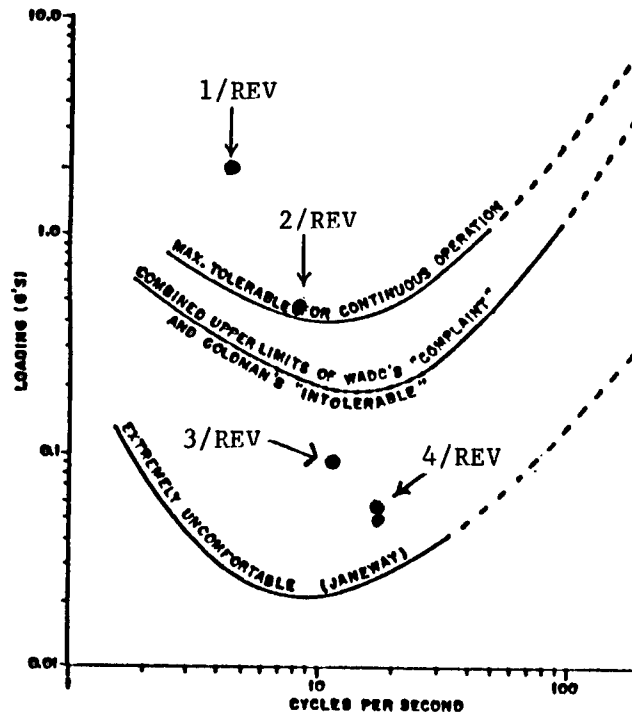
The results are presented in figure II.6.

Figure II.6,Cockpit Vibration Levels

For the undamaged rotor we have one data point in figure II.6. This occurs at a forced frequency ratio of 4.0 because it is the first integer multiple of the number of blades.This means that, for an undamaged rotor, the rotor is in a balanced condition and only frequency ratios that are integer multiples of the number of blades are transmitted into the fuselage. Thus, only ratios equal to 4,8,12 etc. can be transmitted.

In the case of the undamaged Black Hawk, only the frequency ratio of 4.0 had enough amplitude to be significant, the 8.0 and higher frequencies are very small.The 4.0 point is shown in figure II.6 and it is the higher of the two points shown.The other 4.0 point is the vibration level for the damaged rotor.This is not unreasonable because the 4.0 vibrations are not balanced out and then released as unbalanced forces like is the case for the 1.0,2.0, and 3.0 frequencies.Also, the results discussed in this paper report a decrease in root shears for the torsion and 3rd flap modes, and a slight increase in the 2nd flap mode. The forced response that produces vibrations is made up of a combination of all modes plus the effect of possible reduction of response due to changes in the natural frequencies.Thus due to dynamic effects it is entirely possible for the damaged rotor to have 4.0 vibrations less than those for the undamaged rotor.

Also shown in figure II.6 are human tolerance lines as reported by the various investigators from reference 3.These lines represent the subjective comments from the people who were subjected to vibrations in the studies. Figure II.6 shows that the undamaged rotor produces cockpit vibration tolerance that is below the extremely uncomfortable line.

3.Results of Human Tolerance Levels and Forced Vibration Levels for a Damaged Rotor

Results for a hole damaged rotor showed very little additional vibration effects, i.e. it was only slightly changed from the undamaged rotor and produces essentially the same vibration point in figure II.6 as is produced for the undamaged blade. Thus the hole damage points for 1.0,2.0,and 3.0 frequency ratios are so small so as not to appear in Figure II.6. The 4.0 point is coincident with that for the undamaged rotor. This is so because even the removal of half of one blade shifted the 4.0 vibration level only slightly; however the important fact to remember is the other frequencies 1.0,2.0,3.0 etc. are unleashed and they produce the bulk of the vibration increase.

When .96 ft of the tip of the blade is removed, it is also found to have an insignificant effect on cockpit vibration levels.These points again were so small that they do not appear in Figure II.6.It took removal of 13.58 ft from the Black Hawk's normal 26.83 ft radius rotor blade to cause the significant vibration levels shown in figure II.6.

When a rotor blade is damaged the rotor is no longer balanced and all the frequency ratios other than those that are just integer multiples of the number of rotor blades per rotor can be transmitted into the cockpit.

Figure II.6 reflects this inbalance of the rotor, showing the vibration levels for frequency ratios of 1.0,2.0,3.0, and 4.0.We see in figure II.6 that the loss of about half of one blade of a rotor set produces vibration levels that cannot be tolerated for the 1.0 & 2.0 frequency ratios, close to the intolerable and complaint line for 3.0, and above the extremely uncomfortable line for the 4.0 ratio.

4. Correlation of the Vibration Computer Code with Accident Reports

In order to validate the vibration computer code, test data is needed. However, conducting tests of this nature is dangerous to life and the loss of aircraft. The only other source of test data is accident reports that have sufficient engineering detail that could serve as correlating data.

BRL has initiated an effort to accumulate accident report data to determine if correlating data of sufficient detail exists or whether controlled damage testing is required.

III. Conclusions

With respect to the natural vibration characteristics of the Chinook we see the following. In torsion, hole damage tends to reduce the natural frequency. The frequency reduction occurs to a greater extent the farther the hole is positioned inboard on the rotor blade. The shear stress increases to a value of 32% for the most inboard holes.

When portions of the blade tip are lost, there is a significant increase in the torsional natural frequency ratio, and a decrease of the shear stress ratio levels. For a .96 ft loss of the blade tip, the inboard stations have stress ratio levels that are virtually unaffected, but the outboard blade sections experience a large reduction, up to 80% at the tip. This reduction however will be offset in general by an unbalanced rotor that will increase vibration levels transmitted throughout the aircraft.

When the pitch link is lost, a severe change in the torional natural frequency occurs. It changes from 4.62 to 1.0, with an accompaning large shear stress ratio increase to about five times normal.

The general trend in blade torsion is that hole damage or the loss of a pitch link causes a decrease in the natural frequency ratio and a corresponding increase in shear stress somewhere on the blade. Loss of a blade's tip section reduces the internal shear stresses on the blade, but there is an accompaning resulting rotor inbalance.

In flap bending, hole damage reduces the flap natural frequency ratio only for the third mode. The third mode frequency ratio decreases as the hole damage moves inboard. The bending stress ratios associated with hole damage increase to 50-60% for all the hole positions in both the second and third modes.

Loss of portions of the blade tip in flap bending increases the natural frquency ratios of both the second and third modes, and the bending stress ratios are reduced toward the tip sections. The stresses of the inboard sections increase however by 40% in the second mode, and 20% for the third mode.

Also, flap bending damage will contribute additional unbalance to the rotor system.

For the forced vibration analysis on the Black Hawk helicopter,it is found that hole damage and the loss of .96 ft of the tip produce small effects on cockpit vibration levels.It takes a significant loss of a rotor blade up to half the blade to produce severe vibrations that a pilot could not tolerate.

There are additional considerations that are not taken into account in this study and require additional analysis. The types of rotor blade damage that are insignificant as far as vibrations are concerned could be significant in the controllability of the aircraft.That is if a pilot could endure a vibration environment, he may not have enough excess capability in his controls to compensate for various types of aircraft damage to continue his mission.

BRL is continuing going forward with test plans and modeling for a comprehensive analysis to evaluate the flight dynamics of a battle damaged helicopter.

References

1.Pestel,Eduard C.,Leckie,Frederick A.,"Matrix Methods in Elasto Mechanics",McGraw Hill,1963,pages 56 & 57.
2.Scarpati,T.,"CH-47 Fiberglass Rotor Blade Preliminary Structural Analysis Report",Boeing Report DAAJ01-72-C-0676(PIG), 1973,pages 19-26.
3.Linder,G.S.,"Effects of Mechanical Vibrations on Humans",North American Aviation,Report ADA950209,1959,pages 9-11.